



Db 78 PEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLD--NH-----KSTEPMBE----- 125

QY 253 DEIDVVPKSPLDVGSSEVGTGSKNDNKLIRQESCLGNSPPFEKESEPESPMDVDSKNS 312

Db 126 ---DVNKSNI-----DVAINSEDDDELVLLEN-----NKMR-----DGEQVQQL 163

QY 313 QDSEADEE---TSPGFDEQE-----DGSSSQANKPSRFQARDADIEF-----RKR 356

Db 164 SQDLFADDOELIEYPGIMKDTTQDITDSEVETAQKMEMIEEADSTFFVGEDSKATKT 223

QY 357 YSTKG-----GEVRLH-----FQEGGESRTGMNDLNKLP-----RK 388

Db 224 VRTSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTG-NLTLPDLNKKVDPRNRY 282

QY 389 ----NISSLNVECRNSKQHGKDKSKITDHLMLRPLKAEEDRRKEQWETKHORTE-----RK 438

Db 283 CTIPNFPASQKLRDNRYGPK-----IVLP-----QRWREFDSRGRRRDSYFYFKRK 330

QY 439 IPKYVPPHLSPDKKWLGTPIBEMRRM-PRCGIRLPLLRPSANHTVTIRVDLLRAGEVPKP 497

Db 331 LDGYLKCYKTTGYFMFVGLLHNWVEFDPDITYKLPAL----- 368

QY 498 PPTHYKDLNDKNHKVMPCEQNLVPVEDENGERTAGSRWELI-QTALLNKFTRPQNLKDA 556

Db 369 --MYKEM-----SELVGREEVLEKFAVARIA-- 394

QY 557 ILKYNVAYSKKWDF TALIDFWDKVLLEAEAAQHLYQSILPDMVKIALCLPNICTQPIPLLK 616

Db 395 -----KTAEDILPERIYR-LVGDV----- 412

QY 617 QKMNHSITMSQEQIASLLANAFCTFPRRNNAKMKSEYSSYPDINFNRLFEGRSRKPDKL 676

Db 413 -----ESATLSHKQCAALVARMFF-----ARPDSPFS-----FCRILSSDKSICVVEKL 455

QY 677 KTLFCYFRRVTEKKPTGLVTF--TRQSLEDF-PEWERCERKPLTRLHVT--YEGTIEENGQ 731

Db 456 KFLFTYFDKMSMDPPDGAVSFRITKMDKDTFNEEWK--DKKLRLSPEVEFFDEMLIEDTA 513

QY 732 GMLQVDPANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTFVLDHNECLIITGTEQYSE 791

Db 514 LCTQVDFAHEHLGGVNLHGVSQEEIRFLMCPMMVGMVGLCEKMKQLEAISIVGAYVFS 573

QY 792 YTGyAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFR-----YLDQFVPEK 840

Db 574 YTGyGHTLKWAELOPNHSRQNTNEFRDRFGRRLRVETIAIDAILFKGSKLDCQTEQLNKAN 633

QY 841 MRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYF 900

Db 634 IIREMKKASIGFMSQGPKEFTNI-PIVTGWWGCGAFNGDKPLKFLIIQVIAAGVADRPLHFC 692

QY 901 TRGDSELMRDIYSMHIFLTERKLTVG 926

Db 693 SFGEPELAAKCKKIIERMKQKDVTLG 718

RESULT 7

US-10-424-599-184988

; Sequence 184988, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 184988

; LENGTH: 200

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(200)

; OTHER INFORMATION: unsure at all Xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_138059C.1.pep

US-10-424-599-184988

Query Match 4.2%; Score 217.5; DB 12; Length 200;

Best Local Similarity 29.9%; Pred. No. 6.2e-09;

Matches 53; Conservative 27; Mismatches 68; Indels 29; Gaps 4;

QY 625 MSQEQIASLLANAFCTFPRRNNAKMKSEYSSYP-----DINFNRLFEGRSRKPDK 675

Db 33 LTQETSAPPPSRTLFCLEP-----VSDRPGIHLPMINSDVSSGSLYEDYSQKPEST 83

QY 676 LKTLFCYFRRVTEKKPTGLVTFTRQSLE-----DPPE---WERCEKPLTRLHVYEG 724

Db 84 TGRIAHYQORISSEMPKGIVSFERKVLFPFKNDSTHISYPDANFWSTSAIPLCRFEVHSSG 143

QY 725 TIEENGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTFVLDHNECL 781

Db 144 LIEDQSSGAEEVDLANKYLGGALGRGCVQEEIRFMVSPLEAGMLFLPAMADNEAI 200

RESULT 8

US-10-424-599-156445

; Sequence 156445, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 156445

; LENGTH: 180

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_1122291C.1.pep

US-10-424-599-156445

Query Match 3.6%; Score 188; DB 12; Length 180;

Best Local Similarity 30.1%; Pred. No. 1.2e-06;

Matches 49; Conservative 23; Mismatches 41; Indels 50; Gaps 4;

QY 844 ELNKAYCGFL-----RP----- 855

Db 7 EINKAFCGFLYQCKYQYQKILQENGCTSAIFYAATSTMETDEGEISNKHITNSQNDYH 66

QY 856 GVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERD-VVYFTFGDSELMRDIYSM 914

Db 67 GMDQGNINIGVATGNWCGGAFGGDPEVKTIQWLAASQALRPFIAYTTFG-LEALQSLDEV 125

QY 915 HIFLTERKLTVGDVYKLLRLRYNNEECRNCSTPGPDIKLYPFIY 957

Db 126 AHWILSORWTVGDLNMLIEYSINRSKGETNVGFLQWLLPSIY 168

RESULT 9

US-10-424-599-233915

; Sequence 233915, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K





Db 521 SPSIESD-----PEGFEISPEKIIIEVQVYKLP-TAVSLYSPTDEQ----- 560

QY 540 QTALLNKFTRPQNLKDALIKYNVAYSKKWDFTL--IDFWDKVLLEAE-----AQHLY 590

Db 561 --SIMQKEGSQALKSABEYEMMHKTHKYKAFPAANERDEVFEKEPLYGGMLIEDYIY 618

QY 591 QSILPD 596

Db 619 ESLVED 624

RESULT 11

US-09-973-451-11

; Sequence 11, Application US/09973451

; Patent No. US20020132328A1

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE

; TITLE OF INVENTION: (PARG) ENZYMES,

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/973,451

; CURRENT FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: US/09/302,812

; PRIOR FILING DATE: 1999-04-30

; PRIOR APPLICATION NUMBER: 60/083,768

; PRIOR FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 11

; LENGTH: 31

; TYPE: PRT

; ORGANISM: Bos taurus

; FEATURE:

US-09-973-451-11

Query Match 3.2%; Score 167; DB 9; Length 31;

Best Local Similarity 100.0%; Pred. No. 3.6e-06;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 770 LFTVLDHNECLIIITGTEQYSEYTGAEYR 800

Db 1 LFTVLDHNECLIIITGTEQYSEYTGAEYR 31

RESULT 12

US-10-221-278-197

; Sequence 197, Application US/10221278

; Publication No. US20040034208A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: No. US20040034208A1el Nucleic Acids and Polypeptides

; FILE REFERENCE: 21272-045

; CURRENT APPLICATION NUMBER: US/10/221,278

; CURRENT FILING DATE: 2002-09-06

; PRIOR APPLICATION NUMBER: 09/693,267

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 09/665,363

; PRIOR FILING DATE: 2000-09-19

; PRIOR APPLICATION NUMBER: 09/616,847

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 09/596,193

; PRIOR FILING DATE: 2000-06-17

; PRIOR APPLICATION NUMBER: 09/574,454

; PRIOR FILING DATE: 2000-05-19

; PRIOR APPLICATION NUMBER: 09/519,705

; PRIOR FILING DATE: 2000-03-07

; NUMBER OF SEQ ID NOS: 752

; SEQ ID NO 197

; LENGTH: 1163

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-221-278-197

Query Match 3.2%; Score 166.5; DB 12; Length 1163;

Best Local Similarity 20.0%; Pred. No. 0.0011;

Matches 161; Conservative 109; Mismatches 299; Indels 235; Gaps 38;

QY 19 ATTSPAASDARS-----FPSRQRRVLDPKDAHVFQFRVPPSPACVPGQAGQH----- 65

Db 186 AVSSLNSSHRSRSHGNDHHSKEHQRSKSPRDPDANW----DSPSRVPFSSGQHSTQSFPPS 241

QY 66 -RGSATSLVFKQKTTITSWMDTKGIKTAE-----SESLSKKNNTNRIESMMSSVOKDNFYQ 120

Db 242 LMSKSNMLQKPTAYVRPMD--GQESMEPKLSSEHYSSQSHGNSMTLKPSS-----K 292

QY 121 HNVEKLVNVSQSLDKS-----LITEKSTQYLNQHTA--AMCKWQNE-----GKH 163

Db 293 AHLTKLKIPSQ-PLDASASGDVSCVDEILKEMTHSWPPLTAIHTPCKTEPSKFPFPTKE 351

QY 164 TEQLLESEPQTVLVPEQFSNANIDRSPQND--HSDTDSEENRDNQQLTTTVKLAN 218

Db 352 SQQSNFGTEQKRYNPBKTNGHQSKMLKDDLLSSSDSDGEQCDK---TMPRSTP 407

QY 219 AKQTTEDEHAREAKSHQKSKSCHPGEDECACQCOQDEIDVVPKSPLSVDVSGEDVGTGSKND 278

Db 408 GSNSEPSHHNSEGADNSRDDSSSHSGSESS-----GSDSESESSSSD 450

QY 279 NKLIRQESCLGNSPPPEKSEPEP-----MDVDN----- 308

Db 451 SE-----ANEPSQASPEPEPPPTNKWQLDNWLNKVNPHKVPASSVDSNIPSSQY 502

QY 309 -----SKNSCQDSEADEETS---PGFDEQ--EDGSSSQTANKPSRFQARDADIEFRK 355

Db 503 KKEGREQGTGNSYTDTSQPKETSSATPGRDSKTIQKSESGRGRQKSPAQS-DSITQRT 561

QY 356 --RYSTKGGEVRLHFQFEG--ESRTGMNDLNAKLPONISSLNVECRNSKQHGKDKSKI 410

Db 562 VGKKQPKKAEKAAAEPRGGLKIESETPV-DLASSMPSSRRHKAATK-GSRKPNIKKESKS 619

QY 411 TDHLMRLPKAEDRRKEQWETKHQRTERKI-----PKYVPPHLS 448

Db 620 SPR-----PTAE-KKKYKSTKSSQKSREIETDTSSSDSDESESLPPSSQTPKYPESNRT 674

QY 449 PDK-----KWLGTPIEEMRRMPCGIRLPLLRPSANHTVTIRVDLLRAGEVP- 495

Db 675 PVKPSVVEEEDSFQRQMFSPWEEKLLS-----PLSEPDYPLIVKIDLNLLTRIPG 728

QY 496 KPFPPTHYKOLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKD 555

Db 729 KP-----YKETEPKGEKKNVPEKHKTREAQKQASEKVS-----NKGKRKHKNED 772

QY 556 AILKYNVAYSKKWDFTLIDFDWKVLEEAQAHL-----YQSILPDMVKIALCLPNICTOP 611

Db 773 ---DNRASESCK-----PKTEDKNSAGHKPSSNRSSKQSAAKEKDLLPS-PAGP 818

QY 612 IPLLKQKMNHSI---TMSQEQIASLLANAFFCTFPRRNAMKSEYSSYPDINFNRLFEGR 668

Db 819 VPSKDPKTEHGSRKRTISQSSSLKSSSNS-----NKETSGSSKNSSSTSKQKKEGK 870

QY 669 SSRKEPKLKTLCYFRRVTEKKPT 692

Db 871 TSSSSKEVK-----EKAPS 884

RESULT 13

US-10-291-172-197

; Sequence 197, Application US/10291172

; Publication No. US20030228584A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: No. US20030228584A1el Nucleic Acids and Polypeptides

; FILE REFERENCE: 21272-045

```

; CURRENT APPLICATION NUMBER: US/10/291,172
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 09/693,267
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/665,363
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/616,847
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 09/596,193
; PRIOR FILING DATE: 2000-06-17
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 197
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-172-197

Query Match      3.2%; Score 166.5; DB 15; Length 1163;
Best Local Similarity 20.0%; Pred. No. 0.0011;
Matches 161; Conservative 109; Mismatches 299; Indels 235; Gaps 38;

QY 19 ATTSPAASDARS-----PPSRQRRVLDPKDAHVQFRVPPSPACVPGQAGQH----- 65
Db 186 AVSSLSNHSRSHGNDHSHKEHQRSKSPRDPDANW----DPSRVFPFSSGQHSQTSPFPS 241
QY 66 -RGSATSLVFKQKTTITSMWMTGKGTAE-----SESLSKNNNNTRIESMMSSVQKDNFYQ 120
Db 242 LMSKSNMLQKPTAYVRPMD--GQESMEPKLSSEHYSSQSHGNSMTLKPSS-----K 292
QY 121 HNVEKLVNVSQSLSDKS-----LTKSTQYLNQHQHTA--AMCKWQNE-----GKH 163
Db 293 AHLTKLKIPSQ-PLDASAGDVSCVDEILKEMTHSWPPPLTAHTPCKTEPSKFPPTKE 351
QY 164 TEQLLESEPQTVTLVPEQFSNANIDRSPQND-----HSDTDSENRDNQQLFTTVKLAN 218
Db 352 SQQSNFGTGEQKRYNPSTKNGHQSKMLKODLKLSSSDSDGEQCDK----TMRSTP 407
QY 219 AKQTTEDEHAREAKSHQKSKSCHPGEDCASCQQDEIDVVKPSPLSDVGSSEVGTGSKND 278
Db 408 GSNSEPSHHNSEGADNSRDSSSHSGSESS-----GSDSESESSSD 450
QY 279 NKLIRQESCLGNSPPFEKESEPEP-----MDVDN----- 308
Db 451 SE-----ANEPSQASPEPEPPTNKWQLDNWLNKVNPHKVPASSVDSNIPSSQGY 502
QY 309 -----SKNSCQDSEADETS---PGFDEQ--EDGSSQTANKPSRFQARDADIERK 355
Db 503 KKEGREGQGTGNSYTDTSQPKETSSATPGRDSKTIQKSESGRGRQKSPAQS-DSTTQRT 561
QY 356 --RYSTKGGEVRLHFQEGG---ESRTGMNDLNKLPGNISSLNVECRNSKHGKDKSKI 410
Db 562 VGKKQPKKAEKAAAEPRGGLKIESETPV-DLASSMPSSRHKAATK-GSRKPNIKKESKS 619
QY 411 TDHLMRLPKAEDRRKEQWETKHQRTERKI-----PKYVPPHLS 448
Db 620 SPR-----PTAE-KKKYKSTKSSQKSRELIETDTSSSDSDESESLPPSSQTPKYPESNRT 674
QY 449 PDK-----KWLGTPIEEMRMNPRCGIRLPLLRPSANHTVIRVDLLRAGEVP- 495
Db 675 PVKPSVVEEDSFRRQRMFSPMEEEKLLS-----PLSEPDDRPLIVKIDLNLLTRIPG 728
QY 496 KPFPPTHYKDLWNKHVMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKETRPNQKLD 555
Db 729 KP-----YKETEPKGEKKNVPEKHHTREAQKQASEKVS-----NKGKRKHKNED 772
QY 556 AILKYNVAYSKWDFDTALIDFDWKVLEEAQAHL-----YQSILPDMVKIALCLPNICTQP 611
Db 773 ---DNRASESKK-----PKTEDKNSAGHKPSSNRESSKQSAAKEKDLLPS-PAGP 818
```

```

QY 612 IPLLKQKMNHSI---TMSQEQIASLLANAFFCTFPRRNAMKSEYSSYPDINFNRLFEGR 668
Db 819 VPSKDPKTEHGSRKRTISQSSSLKSSNS-----NKETSGSKNSSSTSQKKTEGK 870
QY 669 SSRPEKLTLCYFRRRVTEKKPT 692
Db 871 TSSSSKEVK-----EKAPS 884
```

RESULT 14

```

US-10-282-122A-43837
; Sequence 43837, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43837
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-43837
```

```

Query Match      3.2%; Score 165.5; DB 12; Length 1274;
Best Local Similarity 20.3%; Pred. No. 0.0015;
Matches 104; Conservative 75; Mismatches 173; Indels 161; Gaps 20;
```

```

QY 49 RVPSPSPA-----CVPQAGQHR-----GSATSLVFKQKTTITSMWMTGKIKTA 91
Db 203 RIPVSKPSEKVESDKQYDKYVAKTQTSQNKQLEQEKQNDVVKQGTASKSSDENVSSTT 262
QY 92 ESESLDSKENNNTRIESMMSS-----VQK----- 115
Db 263 KSMPNYSKVDNTIKIENIYASQIVVEIRRERERKVLQRRFPKALQKREHKNEEQDAI 322
QY 116 ----DNFYQHNVEKLVNVSQSLSDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTQLESE 171
```

|    |     |   |     |
|----|-----|---|-----|
| Db | 323 | QRAIDEMYAKQAERYVVGDSLSLNDSDLTNDNSTDASQLHTNGI-----E            | 366 |
| QY | 172 | PQTVTLVPEQFSNANIDRSPQNDHSDTDSEENRDNQQLTITVVKLANAKQTEDE--HAR   | 229 |
| Db | 366 | NETVS-----NDENKQASIONEDTNDTHVDESPYEEVSLNQVSTTKQLSDDEVTVSN     | 419 |
| QY | 230 | EAKSHQKCSKSCHPGEDCASCQQDEIDVVPKSPPLSDVGSEDVGTGSKNDNKLIRQESCIG | 289 |
| Db | 420 | VTSQH-----SALQHNVEVNDKDEL-----KNQSRLIADSEEDG                  | 454 |
| QY | 290 | NSPPFEKESEPESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSQIANKPSRFQAR--   | 347 |
| Db | 455 | AT--NKEEYSGSID-DAEFYELNDTEYDEDTTSNIEDTNRNASEMHVDAPKTQEHAV     | 510 |
| QY | 348 | -----DADIEFRKRYSTKGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNV         | 395 |
| Db | 511 | TESQVNNIDKTVDNEIELAPRHK-KDDQTNLNV-----NSLKTNDVN---DGHV---V    | 556 |
| QY | 396 | ECRNSKHQKGDKSKITDHLMLRPLKAEDRRKEQWETKHQRTERKIPKYVPPHLSPDK---- | 451 |
| Db | 557 | EDSSMNEIEKHNABITENV-----QNEAAESE-QNVEEKTIENVNPKQTEKVSTL       | 606 |
| QY | 452 | -----KWLGTPIEEMRRMPR---CGIRLPLLRP                             | 476 |
| Db | 607 | SKRPFNVMTPSDKGRMMDRKKHSHKVNVPKLP                              | 639 |

**RESULT 15**

US-09-815-242-12955  
; Sequence 12955, Application US/09815242

Patent No. US2002006  
GENERAL INFORMATION:

; APPLICANT: Ohlssen. Kari L.

APPLICANT: WALL, Daniel  
APPLICANT: Zyskind, Jud  
APPLICANT: CHISEN, Ralf

APPLICANT: WALL, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant

APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A

```

```

; CURRENT APPLICATION NUMBER: US/09/81
; CURRENT FILING DATE: 2001-03-21

```

; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848  
 ;  
 ; PRIOR FILING DATE: 2000-05-23

;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27

; PRIORITY FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 : PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2000-12-  
 ; PRIOR APPLICATION NUMBER: 6  
 ; PRIOR FILING DATE: 2001-02-

```

; PRIOR FILING D
;
; NUMBER OF SEQ
;
; SOFTWARE: Fast

```

```

; SOFTWARE:
; SEQ ID NO 1
: LENGTH: 3

```

; LENGTH: VIII  
 ; TYPE: PRT  
 ; ORGANISM: STAPH

### Query Match

3.2%; Score 164.5; DB 9; Length 1111;

Best Local Similarity 20.5%; Pred. No. 0.0015;

Matches 105; Conservative 73; Mismatches 174; Indels 161; Gaps 20;

OV 49 RVPSSPA-----CVPQAGQHR-----GSATSLVFKQKTITSWMDTKGIKTA 91

4  
4  
2  
2  
4  
4  
6  
2  
4

"  
—  
  
—  
  
"  
—  
  
"

—

|    |     |   |     |
|----|-----|---|-----|
| Db | 203 | RIPVSKPSEKVESDKQKYDKYVAKTQTSONKQLEQEKQNDSSVVKQGTASKSSDENVSSIT | 263 |
| Qy | 92  | ESESLDSKENNNTRIESMMSS-----VQK-----                            | 115 |
| Db | 263 | KSMPNYSKVDNTIKENIYASQIIVEIRRERERKVLQRRFPFKALQKQREHKNEEQDAI    | 322 |
| Qy | 116 | ----DNFYQHNVKLVNVSQSLSDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESE   | 171 |
| Db | 323 | QRAIDEMYAKAERYGVDSSLNDDSDLTDNSTDASQLHTNGI-----                | 365 |
| Qy | 172 | PQVTVLVPEQFSNANIDRSPONDDHSDTDSSENRDNQOFLTTVKLANAKQTTEDE--HAR  | 229 |
| Db | 366 | NETVS-----NDENKQASIQNEDTNDTHLDESPYNYEEVSLNQVSTTKQLSDDEVTVSN   | 419 |
| Qy | 230 | EAKSHQKCSKSCHPGEDCASCQQDEIDVVPKSPLSDVGSSEDVGTGSKNDNKLIRQESCLG | 289 |
| Db | 420 | VTSQH-----SALQHNVEVNDKDEL-----KNQSRLIADSEEDG                  | 454 |
| Qy | 290 | NSPPFEKESEPESPMDVNSKNSCQDSEADEETSPGFDEQBDGSSQ---TANKPSRF--    | 344 |
| Db | 455 | AT--NKEEYSGSQID-DAEFYELNDTEVDEDTTSNIEDNTRNASEMHVDAPKTOEYAV    | 510 |
| Qy | 345 | -----QARDADIEFRKRYSTKGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNV      | 395 |
| Db | 511 | TESQVNNIDKTVDNEIELAPRHK-KDDQTNL-----SVNSLKTNDVNDNHVVEDSSMN-   | 562 |
| Qy | 396 | ECRNSKQHGKDKSKITDHLMLPLPKAEDRRKEQWETKHQRTERKIPKYVPPHLSPDK---- | 451 |
| Db | 563 | -----EIEKNNAEITENV-----QNEAAESE-QNVEEKTIENTVNPKKQTEKVSTL      | 606 |
| Qy | 452 | -----KWLGTPIEEMRRMPR---CGIRLPLLRP                             | 476 |
| Db | 607 | SKRPFNVMTPSDKKRMMDRKKHKSXVNPVLPKP                             | 639 |

Search completed: May 26, 2004, 19:18:55

Job time : 84.4317 sec



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 26.8489 Seconds  
(without alignments)  
3496.721 Million cell updates/sec

Title: ' US-09-302-812-4  
Perfect score: 5190  
Sequence: 1 MNAGPGCEPTKATRWGAAT.....YHAVESCAETADHSGQRTGT 976

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID  | Description        |
|------------|-------|---------------|--------|--------|--------------------|
| 1          | 511   | 9.8           | 997    | B84726 | probable poly(ADP- |
| 2          | 463   | 8.9           | 726    | T21138 | hypothetical prote |
| 3          | 340   | 6.6           | 364    | A84726 | probable poly(ADP- |
| 4          | 180   | 3.5           | 1298   | I54367 | X-linked nuclear p |
| 5          | 178   | 3.4           | 1359   | T34036 | hypothetical prote |
| 6          | 168   | 3.2           | 3147   | T18674 | hypothetical prote |
| 7          | 166.5 | 3.2           | 1641   | I38614 | helicase II - huma |
| 8          | 165.5 | 3.2           | 1274   | A89959 | hypothetical prote |
| 9          | 165   | 3.2           | 884    | D96730 | unknown protein F5 |
| 10         | 164   | 3.2           | 4717   | T41581 | hypothetical coile |
| 11         | 161   | 3.1           | 1788   | T29043 | hypothetical prote |
| 12         | 159.5 | 3.1           | 1183   | S65236 | probable membrane  |
| 13         | 159   | 3.1           | 406    | S38170 | SRP40 protein - ye |
| 14         | 159   | 3.1           | 1403   | A47328 | natural killer cel |
| 15         | 159   | 3.1           | 2429   | SJHUA  | spectrin alpha cha |
| 16         | 156.5 | 3.0           | 1538   | T29095 | cardiac muscle fac |
| 17         | 155.5 | 3.0           | 1974   | T30010 | hypothetical prote |
| 18         | 154.5 | 3.0           | 646    | S15901 | chromogranin B pre |
| 19         | 154.5 | 3.0           | 3225   | I52300 | giantin - human    |
| 20         | 154.5 | 3.0           | 3259   | A56539 | giantin - human    |
| 21         | 154   | 3.0           | 1938   | JC5421 | smooth muscle myos |
| 22         | 154   | 3.0           | 1972   | T38411 | probable GTPase ac |
| 23         | 153   | 2.9           | 1031   | T38411 | hypothetical prote |
| 24         | 153   | 2.9           | 1071   | E85343 | protein UNC-89 - C |
| 25         | 153   | 2.9           | 6642   | T29757 | BUD4 protein - yea |
| 26         | 152   | 2.9           | 1320   | S57113 | hypothetical prote |
| 27         | 151   | 2.9           | 1272   | C90593 | RAD50 protein - ye |
| 28         | 150.5 | 2.9           | 1312   | BWBVDL | hypothetical coile |
| 29         | 150.5 | 2.9           | 1957   | T38077 |                    |

|    |       |     |      |   |        |                    |
|----|-------|-----|------|---|--------|--------------------|
| 30 | 150.5 | 2.9 | 1972 | 1 | A41604 | myosin heavy chain |
| 31 | 150   | 2.9 | 1206 | 2 | T34021 | protein kinase SK2 |
| 32 | 150   | 2.9 | 2137 | 1 | SOHUB  | spectrin beta chai |
| 33 | 149.5 | 2.9 | 1210 | 2 | I39410 | AF-4 protein, spli |
| 34 | 148.5 | 2.9 | 472  | 2 | T27903 | hypothetical prote |
| 35 | 148.5 | 2.9 | 665  | 2 | B71609 | hypothetical prote |
| 36 | 148.5 | 2.9 | 677  | 1 | CNHUB  | chromogranin B pre |
| 37 | 148.5 | 2.9 | 1928 | 2 | S46773 | myosin heavy chain |
| 38 | 148   | 2.9 | 644  | 2 | S55395 | neurofilament prot |
| 39 | 148   | 2.9 | 1271 | 2 | A45555 | glutamate rich pro |
| 40 | 148   | 2.9 | 1780 | 2 | T17272 | hypothetical prote |
| 41 | 147.5 | 2.8 | 2218 | 2 | B84683 | hypothetical prote |
| 42 | 146.5 | 2.8 | 667  | 2 | A40713 | cylicin I - bovine |
| 43 | 146.5 | 2.8 | 867  | 2 | T21311 | hypothetical prote |
| 44 | 146.5 | 2.8 | 1463 | 2 | T30290 | AAS surface protei |
| 45 | 146.5 | 2.8 | 2649 | 2 | A40937 | bullous pemphigoid |

ALIGNMENTS

RESULT 1

B84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84726  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-997 <STO>  
A;Cross-references: GB:AE002093; NID:G4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g31870  
A;Map position: 2

|                       |                  |  |             |             |
|-----------------------|------------------|--|-------------|-------------|
| Query Match           | 9.8%             | Score 511;   | DB 2;       | Length 997; |
| Best Local Similarity | 31.8%            | Pred. No. 2.4e-23;   |             |             |
| Matches 135;          | Conservative 58; | Mismatches 118;  | Indels 114; | Gaps 12;    |
| QY                    | 577              | WDKVL EEAQAHLVQSILPDMVKIALCLPNI-----CTQPIPLLKQKMNHS              | 622         |             |
| Db                    | 79               | FDELID EKESKRWFDEIIPALASLLQFPSSLLEVFHQNADNIVSGIKTGLRLNSQQAGI     | 138         |             |
| QY                    | 623              | ITMSQEQIASLIANAFFCTFPRRNAMKMKSESSYPDINFN--LFEGRSSRKPEKLTFLF      | 680         |             |
| Db                    | 139              | VFLSQELIGALLACSFCLFPDDNRGAK---HLPVINFHLSLYISYSQSQESKIRCI         | 194         |             |
| QY                    | 681              | CYFRRVTEKKPTGLVTFTRQ--SLEDFFPEWERCEKPLTRLHVTVYEGTIEENGQGMQVDF    | 738         |             |
| Db                    | 195              | HYFERFCSCVPIGIVSFERKITAAPDADFWSKSDVSL-----YQ-----PDNALEVDF       | 242         |             |
| QY                    | 739              | ANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTVLDHNECLIIITGTEQYSEYTGVAET     | 798         |             |
| Db                    | 243              | ANKYLGGSLSRGCVCQVEIRFMINPELIAGMLFLPRMDDNEAIEIVGAERFSCYTGVA       | 302         |             |
| QY                    | 799              | YRWSRSHEDGSRDCCERRCTEIVAIDAL-----HPRRYLDQFVPEKMRRELKAYCGF        | 852         |             |
| Db                    | 303              | FRPAGEYIDKKAMPFKRRRTRIVAIDALCTPKMRHFKDIC-----LLREINKALCGF        | 355         |             |
| QY                    | 853              | LR-----  | 854         |             |
| Db                    | 356              | LNC SKAWEHQNI FMD EGDNEIQLVRNGRDSGLLRRTTETASHRTPLNDVEMNREKPNANLI | 415         |             |
| QY                    | 855              | -----PGVSSSENL--SAVATGNWCGGAFGGDARLKALIQILAAAAAERDVV-YFTFGDSE    | 906         |             |
| Db                    | 416              | RDYFVEGVDNEDHEDDGVATGNWCGGVFGGDPPELKATIQWLAAASQTRRPFI SYITFG-VE  | 474         |             |



RESULT 4  
I54367  
X-linked nuclear protein - human  
C;Species: Homo sapiens (man)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C;Accession: I54367  
R;Gecz, J.; Pollard, H.; Consalez, G.; Villard, L.; Stayton, C.; Millasseau, P.; Khrestov Hum. Mol. Genet. 3, 39-44, 1994  
A;Title: Cloning and expression of the murine homologue of a putative human X-linked nuc A;Reference number: I54367; MUID:94214473; PMID:8162050  
A;Accession: I54367  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1298 <RES>  
A;Cross-references: GB:I34363; NID:G508653; PIDN:AAA20872.1; PID:G530788  
C;Genetics:  
A;Gene: XNP

Query Match 3.5%; Score 180; DB 2; Length 1298;  
Best Local Similarity 19.3%; Pred. No. 0.0066;  
Matches 199; Conservative 160; Mismatches 374; Indels 296; Gaps 53;

QY 75 KQTIITMMDTKGIKTAESSELSKEN--NNTRIESMMSSVQKDNFYQHNVKLVNVSQ 132  
Db 17 ERRNLSSKRNTKEIQSGSSS--DAEESDNNKKKQRTSSKKAVIVKEKKRNSLRTSTK 75  
QY 133 SLDKSLTEKSTQVNLQHQTAAACKWQNEGKHTEQLLESEPTVTLVPEQ---FSNANIDR 189  
Db 76 RKQADITSSSSDIEDDDQNSI---GEGSDEQKI--KPVTENLVLSHTGFCQSSGDE 129  
QY 190 S-----PQNDHSDTDSEENRDNQFLITVK--LANAKQTTTEDEHAREAKSHQCSKS 240  
Db 130 ALSKSVPTVDDDDNDPENRIAKKMLLEIKANLSSDEGSSDDEPEGKRTGKQNE 189  
QY 241 CHPGEDCASCQDEIDVVPKSPSLDVGSSEVDTGSKNDKLIQESCL--GNSPPEKES 298  
Db 190 ENPGDEEAKNQNS-----ESDSDSEDL--RAKIQHRLFGTNWLLSFGESGE-EKKT 238  
QY 299 EPESPMDV---DNKNSQDSEADETSPGDEQEDGSSSQNTANKPSRFQARDADIEFRK 355  
Db 239 KPKEHEKVGKRRNRKVSSEDSDFQESGVSEEV--SESEDEQRPRTRSAKKAEEENQ 296  
QY 356 R-YSTKGGEVRLHFQFEGG-----ESRTGMNDLNAKLPGN 389  
Db 297 RIYKQKKRRRIKQVEDSSSENKSNSEEEEEKKEEEEEEEEEEDEND--DSKSPG- 354  
QY 390 ISSLNVECRNSKQHGK--KDSKITDHLMLPKAEDRRKEQWETKHQRTKIPKYVPHL 447  
Db 355 -----KGRKKIRKILKDDKLRTETQNALKEBERKERRKRVRERETEKLEREVIEIDA 406  
QY 448 SPDKKWLGTPI-----EEMRRMPCGIRLPLLRPSANHTVTIRVDLLRAGEVPKPPPTHY 502  
Db 407 SPTKCPITTKLVLDDEETKE-----PLVQVHRNMVIL-----KP---HQ 444  
QY 503 KD---LWDNKKVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLINKFTRPQNLKDAIL 558  
Db 445 VDGVOFMWD-----CCCESV-----KTKKSPGSGCILAHCMLGK-----TLQVVSF 487  
QY 559 KYNVAYSKKWDF-TALIDFWDKVL-----EEAQAHLQSIPLPDMVKIALCLPNICTQPI 612  
Db 488 LHTVLLCDKLDFTALVGLSSSILAFNMWNEFEK---WQEGLKDDKLE--VSELATVKR 542  
QY 613 PLLKQKNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNLFEGRSSRK 672  
Db 543 P---QERSYMLQRWQEDGGVMIIG-----YEMY-----RNLAQGRNV-K 577  
QY 673 PEKLTLCFYRRRVTEKKPTGLVTFTTRQSLDFPWERCEKPL-----TRLHVTYEGT--- 725  
Db 578 SRKLKEIF--NKALVDPGPDFVVCDEGHILKN--EASAVSKAMNSIRSRRIILTGTPLQ 633  
QY 726 -----IEENGQGMLOQDFANRFV-----GGGVTS----- 749  
Db 634 NNLLEYHCMVNFIKENLLGSIK-EFRNRFINPIQNCQCADSTMVDVRVMKKRAHILYEML 692

QY 750 AGLVQEE-----IRFL-INPELIISRLFTFVLDHNECLITGTGEQYSEYTGAE--TYRW 801  
Db 693 AGCVQRKDYTALTCKFLPKPKHEYVLAVRMTSI-----QCKLYAGAKLFQDFQMLSRITWHPW 748  
QY 802 -----SRSHEDGSEDDCERRCTEIVADALHFRRYL--DQFVPEKMRRELNKAYCGF 852  
Db 749 CLQLDYISKENKGYFDEDSMD---EFIASDSDETSMSLSSDDYTKKKKGKKGK----- 799  
QY 853 LRPGVSSSENLSAVATGN-----WGCGAFGGDARLKALIQILAAAAAERDVVYFTFG 903  
Db 800 -----KDSSSSGSGSDNDVEVIKWNRSRSGGGE-----G 829  
QY 904 DSELMRDIYSMHIFLTERKLTGVGDVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESC 963  
Db 830 NVDETGNPNPSVSLKLEESKAT-----SSSNPSSPAPDW-----YKDFVTIDAD 871  
QY 964 AETADHSGQ 972  
Db 872 AEVLEHSGK 880

RESULT 5  
T34036  
hypothetical protein B0041.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T34036  
R;Fulton, R.; Wohldmann, P.  
submitted to the EMBL Data Library, April 1997  
A;Description: The sequence of C. elegans cosmid B0041.  
A;Reference number: Z21466  
A;Accession: T34036  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1359 <FUL>  
A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7  
A;Experimental source: strain Bristol N2; clone B0041  
C;Genetics:  
A;Gene: CESP:B0041.7  
A;Map position: 1  
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 3.4%; Score 178; DB 2; Length 1359;  
Best Local Similarity 19.1%; Pred. No. 0.0093;  
Matches 147; Conservative 118; Mismatches 279; Indels 224; Gaps 33;

QY 86 KGIKTAESSELSKNNNTRIESMMSSVQKDNFYQHNVKLVNVSQSLDKSLTEKSTQY 145  
Db 81 KSRKRAKSES-ESDESDEEDRKKSKKKVQDKKQKESKKGRTTSSSEDESDEREQ- 138  
QY 146 LNHQHTAAMCKWQNEGKHTEQLLESEPTVTLVPEQFSNANIDRSPQNDHSDTDSEENR 205  
Db 139 -KSKKKSKTKKQTSSESSE--ESEERKVKKSKKKEKSVKKRAETSESEDEKPSK 194  
QY 206 DNQQLTTVKLANAKQTTTEDEHAREAKSHQKCSKSCHPGEDCASCQDEIDVVPKSPLS 265  
Db 195 KSKKGLK--KKAKESESESEDEKEVKSKKKKSKKVKVKESESEDEAPKKTEKRKRSK 252  
QY 266 VGSEDVGTGSKNDNKLIRQESCLGNSPPFEKSESE--ESPMDVNSKNKNSCQDSEADEE 321  
Db 253 TSSESESESEKSDDEE---EE-----EKESPPKPKKKPLAVKKLSSDESESDVE 300  
QY 322 TSP-----GFDEQEDGSSSQNTANKPSRFQARDADIEFFKRYSTKGGEVRLHF 368  
Db 301 VLPQKKKRGAVTLISDSEDEKQKSESEASDVEEKVSKGA---KKQESSE----- 348  
QY 369 QFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHG--KKDSKITDHLMLPKAEDRRKE 426  
Db 349 --SGSDSSEGSITVNRK-----SKKKEKPEKKKKGIIMDSSKLQKETIDAERAERK 400  
QY 427 QWETKHQR-----TERKIPKYVPPHLSPDKKWLGTPIEEMRRMPR 466

```
Db 401 RLEKKQKEFNGIVLEEGEDLTEMLTGTSSQRKLSVV---LDPD-----SSTVDEESKKP- 452
Qy 467 CGIRLPLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVMKPCSEQNLYPVEDE 526
Db 453 -----VEVHNSLVR---ILKPHQAH-----GIQFMYDCAFE 480
Qy 527 NGER--TAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFETALIDFWDKVLBEA 584
Db 481 SLDRLDTEGS-----GGILAHCMGLGKT---LQVITFLHTVL--- 514
Qy 585 EAQHLVQSILPDMVKIALCLP-NICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFP 643
Db 515 ----MHEKIGEKCKRVLVVVPKNVIINWF-----KEFQKWLVDNDEELDTIDVN----- 559
Qy 644 RRNAKMKSEYSSYPDINFNR--LPEGRSSRKPEKLTLCFCYFRRVT-----EKKPTGLV 695
Db 560 -----ELDSYKTIEDRRALKAWHSSKTPSVMIIGYDLFRILTVEDDPKKKKPKNRN 611
Qy 696 TTFRQSLDEF-----PEWERCE-----KPLTRLHVITYEGTIEENG 730
Db 612 RRLKAKEDFRKYLQNPQPMVMVCDDEAHKLNDDSAKSCVMVKILTTRRICLTGTPLQNN 671
Qy 731 -----QGML--QVDFANRFVGGVTSAGLVQE-----EIRFL 760
Db 672 LMEYHVMVNFVKPGLLGTCTEFANRFV--NIINRGRTKDASPLEVSFM 717

RESULT 6
T18674
hypothetical protein T04F3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T18674; T24464
R:White, S.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19004
A:Accession: T18674
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3147 <W1L>
A:Cross-references: EMBL:Z74026; PIDN:CAA98419.1; GSPDB:GN00023; CESP:T04F3.1
A:Experimental source: clone B0240
R:Kershaw, J.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19894
A:Accession: T24464
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3147 <W12>
A:Cross-references: EMBL:Z72513; PIDN:CAA96672.1; GSPDB:GN00023; CESP:T04F3.1
A:Experimental source: clone T04F3
C:Genetics:
A:Gene: CESP:T04F3.1
A:Map position: 5
A:Introns: 338/3; 417/3; 1957/2; 2358/3; 2400/1; 2529/2; 2585/1; 2700/3; 2744/1; 2877/3;

Query Match
Best Local Similarity 3.2%; Score 168; DB 2; Length 3147;
Matches 155; Conservative 114; Mismatches 322; Indels 208; Gaps 30;

Qy 64 QHRGSATSLVFK-----QKTITSWMDTKGIKTAESLDSKENNTR 105
Db 1042 EHRTSAVNIDLEKVFHIGSSKKPKNDDEKIRRGIAEFERTKQKEAQRSTVIETQSNSR 1101
Qy 106 I--ESMMSSVQKDNFYQHNVKLVNVSQISLDSLSTEKSTQYLNQHQTAAMCKWQNEGKH 163
Db 1102 IFEESSISMDVFNNSLHNESQVSEITASDPDLVLTSTTFHNVIEE---KIDDDVTK 1157
Qy 164 TEQLLESEPOTVLVPEQFSNANIDRSPQNDHSDTDSEE---NRDNQQLTTVKL----- 216
Db 1158 TDSNVVEEKEQVRLRIDEFKRPTEEQNLQKPELTKEEYSVRMENRTSAVSIDLKVF 1217
Qy 217 -ANAKQTT-----EDEHAREAKSHQK--CSKSCHPGE---DCASCQQQDEI- 255
```

```
Db 1218 DQSSKETTVSNETDEKIKRGIAEFERSKOEKQVORSGVAETSHSGKHIFDESISMDDVF 1277
Qy 256 -----DVVPKSPLSDVGSE-----DVGTGSKNDNKLIRQESCLGNP 292
Db 1278 NTSQYKSDKELSSPERTVEPEVSTATMNLNIIIFASGIATRENTDVLSEERIQRVE 1337
Qy 293 PFKESEPE-ESPMVDVNSKNKSCQDSEADEETSPGFDEQEDG-----SSSQANKPFRFQA 346
Db 1338 EFKTTTENLEIQKEVVLTKBEVDNSDVKEHRTSAVNIDLDVFIQRSSKHPEDEDEKI 1397
Qy 347 RDADIEFRKRYSTKGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKK 406
Db 1398 RRGIAEFERTKQEK--EAQRSAVIETQSNNKHIFD-----KSNISM--DEVFNESQNGQK 1448
Qy 407 DSKITDHLMLRPLKAEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRMFR 466
Db 1449 DSSNID-----MKETDMPKERDDQRYVDVH--RDKK---PFENGFEFPT 1488
Qy 467 CGIRLPLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVMKPCSEQ----- 518
Db 1489 FNGSKISNEPKQISITINLD-----NVFPTTEPKLVAEDNCEIEAEERIRKRIKQ 1540
Qy 519 -----NLYPVEDENGE---RTAGSRWELIQTALLNKFTRPON-----LKDA 556
Db 1541 FERTTGEQEIILKNSEPAEDETSDEKKHRTAAVSIIDLKVFVQGTAKKPENDFEDEKIKRG 1600
Qy 557 ILKY-----NVAYSKKWDFETALIDFWDKVLEEAQHLVQS----- 592
Db 1601 IAEFERSKOEKQVORSGVAETSHSGKHIFDESISMDDVENTSOK---YKSEKLSPTPER 1657
Qy 593 -ILPDMVKIALCLPNICTOPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRR----- 645
Db 1658 TVEPEVSTATMNLNIIIFASGIATREK-NTDVLEEEERIQRVEE-----FKKTENLEI 1711
Qy 646 -----NAKMKSEYSSYPDINFNRLFEGRSSRKP-----EKLKTLFCYFRRVT 687
Db 1712 QKEVVLTKEGDSDVKDHKASAVNIDLDVFIQRSSKHPEDEDEKIRRGIAEFERTK 1771
Qy 688 EKQPT--GLVTFTRQSLD 704
Db 1772 QEKEAQRSTVIETQYSSKD 1790

RESULT 7
I38614
helicase II - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
C:Accession: I38614
R:Stayton, C.L.; Dabovic, B.; Gulisano, M.; Gecz, J.; Broccoli, V.; Giovanazzi, S.; Boss
Hum. Mol. Genet. 3, 1957-1964, 1994
A:Title: Cloning and characterization of a new human Xq13 gene, encoding a putative heli
A:Reference number: I38614; MUID:95179111; PMID:7874112
A:Accession: I38614
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1641 <RES>
A:Cross-references: EMBL:U09820; NID:g606832; PIDN:AAC50069.1; PID:g606833
C:Genetics:
A:Gene: RAD54L

Query Match
Best Local Similarity 3.2%; Score 166.5; DB 2; Length 1641;
Matches 206; Conservative 157; Mismatches 366; Indels 325; Gaps 56;

Qy 75 KQKITSWMDTKGIKTAESLDSKEN--NNTRIESMMSSVQKDNFYQHNVKLVNVSQ 132
Db 277 ERNLSKKRNTKEIQSGSSSS--DAEESSEDNKKKKQRTSSKKKAVIVKKNRSLRTSTK 335
Qy 133 SLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTQELLESEPOTVLVPEQ---FSNANIDR 189
Db 336 RKQADITSSSSSDIEDDDQNSI---GEGSSDEQKI--KPVTENLVLSSTHTGFCQSSGDE 389
```



```

190 QY S-----PQNDHSDTDSENRDNQQFLTVK--LANAKQTTTEDEHAREAKSHQCSKS 240
      :      ||| ||| :      :      :      :      :      :      :      :
390 Db ALSKSVPTVDDDDDDNDPENRIAKKMLLEIKANLSSDEDEGSSDDEPEEGKRTGKQNE 449
      :      :      :      :      :      :      :      :      :      :
241 QY CHPGEDCACQODEIDVVPKSPLSDVGSSEVGTGSKND--NKLIRQESCLGNSPP-FEK 296
      :||:| :      :      :      :      :      :      :      :      :
450 Db ENPGDEEAKNVNS-----ESDSDSEE--SKPRYRHRLLRHKLTVSDGESGEEK 497
      :      :      :      :      :      :      :      :      :      :
297 QY ESEPESPMDV---DNSKNSQDSEADEETSPPGDEQEDGSSSQTANKPSRFQARDADIEF 353
      :||:| :      :      :      :      :      :      :      :      :
498 Db KTKPKHEKVGKRRNRKVSSEDSDFQESGVSEEV--SESEDEQRPRTRSAKKALEE 555
      :      :      :      :      :      :      :      :      :      :
354 QY RKR-YSTKGGEVRLHFQFEGG-----ESRTGMNDLNAKL P 387
      :      :      :      :      :      :      :      :      :      :
556 Db NORSYKQKKRRRIKQVEDSSSENKNSSEEEEEKEEEEEEEEEDEEDND-DSKSP 614
      :      :      :      :      :      :      :      :      :      :
388 QY GNISSINVECRNSKQHGK--KDSKI-TDHLMLPLKAEADRRKEQWETKHORTERKIPKYVP 444
      :      :      :      :      :      :      :      :      :      :
615 Db G-----KGRKKIRKILKDDKLRKTETQNALKEEERRKRIARERER--EKLREVIE 663
      :      :      :      :      :      :      :      :      :      :
445 QY -PHLSPDKKWLGTPI-----EEMRRMPCGIRLPLLRPSANHTVTIRVDLLRAGEVPKPF 498
      ||| :      :      :      :      :      :      :      :      :
664 Db IEDASPTKCPITTKLVLDEDEETKE-----PLVQVHRNMVIKL-----XP- 703
      :      :      :      :      :      :      :      :      :      :
499 QY PTHYKD---LWDNKHVKMPCSEONLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLK 554
      :      :      :      :      :      :      :      :      :      :
704 Db --HQVDGVQFMWD-----CCCESV-----KTKKSPGSGCIIAHCMGLCK-----TLQ 744
      :      :      :      :      :      :      :      :      :      :
555 QY DAILKYNVAYSKKWDF-TALIDFWDKVL-----EEABAQHLYSILPDMVKIALCLPNIC 608
      :      :      :      :      :      :      :      :      :      :
745 Db VVSFLHTVLLCDKLDFTALVGLSSSILAFNWNNEFEK---WQEGLKODEKLE--VSELA 799
      :      :      :      :      :      :      :      :      :      :
609 QY TQIPILLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 668
      :      :      :      :      :      :      :      :      :      :
800 Db TVKRP---QERSYMLQWQEDGGVMIIG-----YEMY-----RNLAQGR 835
      :      :      :      :      :      :      :      :      :      :
669 QY SSRKPEKLTLCFYFRRYTEKKPTGLVFTFRQSLDFPEWERCEKPL---TRLHVTYEG 724
      :      :      :      :      :      :      :      :      :      :
836 Db NV-KSRKLEIF--NKALVDPPGPDFVVCDEGHILKN--EASAVSKAMNSIRRRRIILTG 890
      :      :      :      :      :      :      :      :      :      :
725 QY T-----IEENGQGMLOQVDFANFV-----GGGVTs-----749
      :      :      :      :      :      :      :      :      :      :
891 Db TPLQNNLIEYHCVNFIKENLLGSIK-EFRNRFNFIQNGQCADSTMVDVRVMKKRAHIL 949
      :      :      :      :      :      :      :      :      :      :
750 QY ----AGLVQEE---IRPL-----INPELIISRLPTEVLHDNECLITGTEQYSE 791
      :||:| :      :      :      :      :      :      :      :      :
950 Db YEMLAGCVQRKOYTALTAKPLPPKHEYVLAVRMTSIOCKLYQYYLDH-----LTGVGNNS 1004
      :      :      :      :      :      :      :      :      :      :
792 QY -----YTGYAETYRW-----SRSHEDGSEERDDCERRCTEIVAIDALHF 829
      :      :      :      :      :      :      :      :      :      :
1005 Db GGRGKAGAKLFQDFQMLSRITWHPWCLQLDYISKENKGYFDEDSMD---EFIASDSDET 1060
      :      :      :      :      :      :      :      :      :      :
830 QY RRYL--DQFVPEKMRRELNKAYCGFLRPGVSSSENLSAVATGN-----WGCGAFGGD 878
      :      :      :      :      :      :      :      :      :      :
1061 Db SMSLSSDDYTKKKKKGKKGK-----KDSSSSGSGSDNDVEVIKWNRSRGGG 1108
      :      :      :      :      :      :      :      :      :      :
879 QY ARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNE 938
      :      :      :      :      :      :      :      :      :      :
1109 Db E-----GNVDETGNNPVSVLKLEESKAT-----1131
      :      :      :      :      :      :      :      :      :      :
939 QY ECRNCSTPGPDIKLYPIYHAVESCAETADHSQG 972
      :      :      :      :      :      :      :      :      :      :
1132 Db SSSNPSSPAPDW---YKDFVTDADAEVLEHSGK 1161
      :      :      :      :      :      :      :      :      :      :

```

RESULT 8  
A89959  
hypothetical protein SA1562 [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: A89959  
C;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: A89959  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1274 <KUR>  
A;Cross-references: GB:BA000018; FID:g13701536; PIDN:BAB42830.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: SA1562

|    | Query Match           | 3.2%;   | Score 165.5;                         | DB 2;                      | Length 1274; |
|----|-----------------------|---|--------------------------------------|----------------------------|--------------|
|    | Best Local Similarity | 20.3%;  | Pred. No. 0.05;                      |                            |              |
|    | Matches 104;          | Conservative 75;  | Mismatches 173;                      | Indels 161;                | Gaps 20;     |
| QY | 49                    | RVPSPSPA-----   | CVPGQAGQHR----                       | GSATSLVFKQKTIITSWMDTKGIKTA | 91           |
| Db | 203                   | RIPVSKPSEKVESDKQYDKYVAKTQTSQNKQLEQEKQNDSSVVKQTASKSSDENVSSTT   | 262                                  |                            |              |
| QY | 92                    | ESESLDSKENNNTRIEBSMMSS-----                                   | VQK-----                             |                            | 115          |
| Db | 263                   | KSPNYSKVDNTIKIENIYASQIVEEIRRRERKRVLQKRRFKKALQOKREHKNEEQDAI    | 322                                  |                            |              |
| QY | 116                   | ----DNFYQHNVEKLVNVSQLSLDKSLTEKSTQYLNQHOTAAAMCKWONEGKHTQELLESE | 171                                  |                            |              |
| Db | 323                   | QRAIDEMYAKQAEYVGDSSLNDDSDLTDNSTDASQLHTNCI-----                | E 365                                |                            |              |
| QY | 172                   | PQTVTLVPEQFSNANIDRSPQNDDHSDTDSEENRDNQFLTTVKLANAKQTTETE--      | HAR 229                              |                            |              |
| Db | 366                   | NETVS-----NDENKQASIQNEDTNDTHVDESPYNEEVSINQVSTTKQLSDDEVTVSN    | 419                                  |                            |              |
| QY | 230                   | EAKSHQCKSKSCHPGEDCASCQQDEIDVVPKSPLSVDGSEDVGTGSKNDNKLIRQESCLG  | 289                                  |                            |              |
| Db | 420                   | VTSQHQ-----SALQHNVEVNDKDEL-----                               | KNQSRLIADSEEDG 454                   |                            |              |
| QY | 290                   | NSPPFEKESEPESPMDVDNSKNCSQDSEADEETS                            | PGFDEQEDGSSSQTANKPSRFQAR--           | 347                        |              |
| Db | 455                   | AT--NKEEYSGSID-DAEFYELNDTEVDEDTTSNIEDNTNRNASEMHVDAPKTQEHAV    | 510                                  |                            |              |
| QY | 348                   | -----DADIEFRKRYSTKGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSNLN         | 395                                  |                            |              |
| Db | 511                   | TESQVNNIDKTVDNEIELAPRHK-KDDQTNLNV-----                        | NSLKTNDVN---DGHV-----V 556           |                            |              |
| QY | 396                   | ECRNSKQHGKKDSKITDHLMLPLPKAEDRRKEQWETKHQTERKIPKYVPPHLS         | SPDK----                             | 451                        |              |
| Db | 557                   | EDSSMNEIEKHAETENV-----  | QNEAAESE-QNVZEKTIENVNPKQTEKVSITL 606 |                            |              |
| QY | 452                   | -----KWLGTPIEEMRRMPR----                                      | CGIRLPLLRP 476                       |                            |              |
| Db | 607                   | SKRPENVMTPSDKKRMMDRKKHSHKSVNVPELKP                            | 639                                  |                            |              |

RESULT 9  
D96730  
unknown protein F5A18.20 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: D96730  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: D96730

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-884 <STO>  
A;Cross-references: GB:AE005173; NID:G6453896; PIDN:AAF09079.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F5A18.20  
A;Map position: 1

Query Match 3.2%; Score 165; DB 2; Length 884;  
Best Local Similarity 20.6%; Pred. No. 0.032;  
Matches 116; Conservative 79; Mismatches 197; Indels 170; Gaps 25;

QY 16 WGAATTS-----PAASDARSFSPRRRV-LDPKDAHVQ---FRVPPSPACV--PGQAG- 63  
Db 317 WQHTTSGVVVPIPPSSAQSIPOHDSMAIPVSGHIMPPYGRFPFPNPQVGPYPYAFGT 376  
QY 64 -----QHRGSATSLVFKQKTTITSWMDTKGIKT-----AESLSD--- 97  
Db 377 KPPLHPVAAFMDDSYAASSVPPKAPVPVNLKBELLKKKADLGRFSSGRFEERESMDDDV 436  
QY 98 -----SKENN-----NTRIESMMSSV---QKDNFYQH 121  
Db 437 LYKPPTKADQPKKSFSPNSSDEEEEDMDAARTTEINMEIKRILTEVLLKVTDELDFDE 496  
QY 122 NVEKLVNVSQ-LSLDKSLTEK---STOYLNQ---HOTAAMCKWONEGKHTEQLLESEPQ 173  
Db 497 IATKVINEDEAIPKDDSVQHNLKSSLLSTADPLHKASAKILVSVEGANTKASSGSPAD 556  
QY 174 TVTLVPEQFSNANIDRSPQNDHSDTDSEEN-----RDN-QQFLTTVKLANAKQTT 223  
Db 557 VLGLA---SYASDDDDADTDAAADANADENGVELGVSRRHNVSQPSTEKL-----P 606  
QY 224 EDEHAREAKSHQKSKSCHPGEDECASCQD-----EIDVVPKSPSLDV---GSED 270  
Db 607 DPEAMASAKLDPAGVNVANSKNSKSGLEDYSQMPGSTRKDDAEGSTKISDVSSAGLDD 666  
QY 271 VGTGSNDNKLIRQESCLGNSPPFEKESEPEPMDVDNSKNSCQDSEADEETSFGFEQE 330  
Db 667 DTSGSRKEH-----PDRDSDKDAILDDEPHVKNS-----GVKS 699  
QY 331 DGSSSQTANKPSRFQARDADIEFRKR-YSTKGGEVRLHFQFEGGESRTGMNDLNAKLPGN 389  
Db 700 DCNLRQDSNKPYGKDLSDVSTDRSRIVETKGGKEKGDQ---NDSKDRMKENDLKSAEK 756  
QY 390 ISSLNVECRNSKHGKKDSKITDHLMLRPLKAEDE-----RKEQWETKHQRTERKIPKYV 443  
Db 757 VKGVESNKKSTDPHVKDSRDVERPHRTNSKEDRGKRKEKEKEERSRHRRAENS----- 811  
QY 444 PPHLSPDKKWLGTPIEMRRMP 465  
Db 812 ----SKDKR-----RRSP 820

RESULT 10  
T41581  
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-Aug-2003  
C;Accession: T41581  
R;Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1998  
A;Reference number: Z22002  
A;Accession: T41581  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-4717 <MUR>  
A;Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08  
A;Experimental source: strain 972h(-)  
C;Genetics:  
A;Gene: SPDB:SPCC737.08  
A;Map position: 3  
C;Superfamily: midasin (AAA ATPase with von Willebrand factor type A (vWA) domain)

Query Match 3.2%; Score 164; DB 2; Length 4717;  
Best Local Similarity 19.6%; Pred. No. 0.38;  
Matches 187; Conservative 143; Mismatches 326; Indels 296; Gaps 51;  
QY 90 TAESESLDSKENNTRIESMMSSVQKDNFYQHN-----VEKLVNVSQSLDKSL 138  
Db 3942 TANQSDLD--ESEARELESMDMGVTKDSVSVSENNSDSEENQDLDEEVNDIPEDLSNLS 3999  
QY 139 TEKSTQYLNQHOTAAMCKWONEGKHTEQLLESEPOTVTLVPEQFSNANIDRSPQNDH-- 196  
Db 4000 NEK-----LWDEPNE--EDLLETEQKS---NEQSAANNESDLVSKEDDNKA 4040  
QY 197 -SDTDSSENRDNQQFLTTVKLANAKQTTTEDEHAREAKSHQKSKSCHPGEDECASCQD-- 253  
Db 4041 LEDKDRQEKEDEEEMSDVGI-----DDEIQPDIQENN--SQPPPENEDHLDLPEDLK 4091  
QY 254 ----EIDVVPKSPSLSDVGSSEVGTGSKNDNKLIRQESCLGNSPP---FEKESEPESPMDV 306  
Db 4092 LDEKEDGVSKDSLEDMDME---AADENK---EADAEKDEPMQDFEDPLEENNLTDE 4143  
QY 307 DNSKNSCQDSEADEE--TSPGFDEQ-----EDGSSSQA-----NKP-- 341  
Db 4144 DIQDDFSDLAEDDEKMNEDGFENNVQENNEESTEDGVKSDEBELLEQGEVPEDQADNHPKM 4203  
QY 342 ---SRFQARDADIEFRKRYSTKGGEVRLHFQF--EGGESRTGMNDLNAKLPGNISLNV 396  
Db 4204 DAKSTFASAEADEE-----NTDKGIVGENEELGEEDGAESGVRG-NGTADGEFSSAEQV 4257  
QY 397 CRNSKHGKKDS-----KITDHLMLRPLKAEDERRKEQW---ETKHQRTERKIPKYV 443  
Db 4258 QKGEDTSTPKEAMSEADROYQSLGDHLREWQQA--NRIHEWEDLTESQSQAFFDSEFMHV 4315  
QY 444 PPHLSPDKKWLGT---PIEEMRMRCGIRLPLLRPSANHTVIRVDLLRAGEVPKPF 499  
Db 4316 KEDEEEDLQALGNAEKDQIKSIDRDESAN-----QNPDSMNSTNIAED--EADEVGD--- 4365  
QY 500 THYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI-QTALLNKFTRPQNLKDAIL 558  
Db 4366 --KQLQDG-----QDISDIK-QTGEDTLPTFGSINQSEKVFELSEDEDEIEDEL 4412  
QY 559 KYNVAYSKKMDFTALI-----DFWDKVLSEAEAAQHLYSILPDMVKIALCLPNICTQPI 612  
Db 4413 DYNV---KITNLPAMPIDEARDLWNK-HEDSTKQ-----LSIELC----- 4449  
QY 613 PLLKQKMHNSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINENRLEFGRSSR- 671  
Db 4450 -----EQLRLILE-----PTLATKMQGDFRTGKRLNMKRIIPIYASQF 4487  
QY 672 KPEKLTLCYFRRRVTEKKPTGLVTFTFQSLQEDFPWEWERCEKPLTLRLHVTYEGTIEENGQ 731  
Db 4488 KKDKI-----WMRRVKPSKRTYQVMI---SIDD-----SKSMSESGS 4521  
QY 732 GMLQVDFANRFVGGGVTSAGLVQE-----EIRFLINPELI--ISRLETEVLDHNECLIIIT 784  
Db 4522 TVLALETL-----ALVTKALSILLEVGQIAVMKFGQPELHHPFDKQFSSE-----S 4567  
QY 785 GTEQYSEYTGVAETRWRSRSHEDGSRDDCERRCTEIVAIDALHFRYLDQFVPEKMRRE 844  
Db 4568 GVQMFSHFT-----FEQSNNTNVLALADASMKCFN-YANTASHRSNSD-----IRQ 4612  
QY 845 LNKAYCGFLPFPVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAARDVVYFTFGD 904  
Db 4613 LEI-----IISDGICEDH-----DSIRKLRRRAQBEKVMIVFVILD 4648  
QY 905 SELMRDIYSMHIFLTERKLTGVGVYKLLRLRYNNEECRNCSTPGPDIKLYPFI 956  
Db 4649 N-----VNTQKSSILDIKKV---YYDTK----EDGTMDLKIQPYI 4682

RESULT 11  
T29043  
hypothetical protein B0228.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T29043  
R;Leimbach, D.  
submitted to the EMBL Data Library, March 1995  
A;Description: The sequence of C. elegans cosmid B0228.  
A;Reference number: Z18324  
A;Accession: T29043  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1788 <LEI>  
A;Cross-references: EMBL:U23168; PIDN:AAC38806.1; CESP:B0228.2  
A;Experimental source: strain Bristol N2  
C;Genetics:  
A;Gene: CESP:B0228.2  
A;Introns: 1456/2; 1482/3; 1516/2; 1551/3; 1595/3; 1646/1; 1671/1; 1716/2; 1749/3  
  
Query Match 3.1%; Score 161; DB 2; Length 1788;  
Best Local Similarity 19.8%; Pred. No. 0.15;  
Matches 213; Conservative 126; Mismatches 411; Indels 328; Gaps 45;  
  
QY 10 CTKATRWGAATTSAAADARSFPQRQRRLDPKDAHVFVRVPPSPACVPQAGQHRGSA 69  
DB 808 CTKE-----ASTSAAALASGS-----KSESCKIVRLASNLGHP 841  
  
QY 70 TSLVFKQKTITSWMDTKGIKTAESSELDSEKNNNNTRIESMMSSVQKDNFYQHNV----- 124  
DB 842 TSLVL-----CESSSVQETNNVH-----YORDEHHEHISETSYP 876  
  
QY 125 -----KLNVNSQLSLDKSLTEKSTQYLNQHQTAAAMCKWQNEGKHTEQLLESEPQT 174  
DB 877 RDGGKFTLDTKASTANEVRIDKLEKSDREL---ETEIKTIVRNEGEPPVEIFVSATEES 933  
  
QY 175 VTLVPEQFSNAN-----IDRSPQNDHSDTDSEENRDNQFLTTVKLANAKQTTEDEH 227  
DB 934 AAGVTTSLSRANPFESANILLTSPNKGEPAYSRVTESSE---LT-----ETNNVQ 980  
  
QY 228 AREAKSHQCKSKSCHPGEDCASCQQDEIDVVPKSPLSVDVSGEDVGTGSKNDNKLIRQESC 287  
DB 981 LRREEHQETEKI-----IETA-----ANGSSLLRAGA-DEKFADIEAK 1020  
  
QY 288 LGNSPPFEKESEPEPMDVDNNSKNSQDSEADEETSPGFDE--QEDGSSSQTANKPSRFQ 345  
DB 1021 LGKDAQFE---SAQTIRQIGNEDKTNLSIGASQETSVTFTDVTQCNKSSCEETSITKVAK 1077  
  
QY 346 ARDADIEFRKRYSTKGGE--VRLHFQFEGGESRTGMNDLNKLPGNITSSLNVECRNSKQH 403  
DB 1078 NIEPHVIFR---STEASDMAGVIHYTLRSSEKVEETEEK-NIARNGGSATFSCFASGD- 1132  
  
QY 404 GKOSKITDHLMLPKAEDRRKEQWETKHQTERKIPKYVPPHLSPDKKWLGTPIEEMRR 463  
DB 1133 -ESPDSVSAFLTRQPQEE-----TTE---KLFTPTMFDFFIKFNSTAAEEFAV 1175  
  
QY 464 MPRCGIRLPLLRPSANHTVTIRVDLLRAGEVPKPPFTHYKDLWDNKHVKMPCSEQNLYPV 523  
DB 1176 W-----NTTIFRRKD--NEGEVEKIFNTS----- 1197  
  
QY 524 EDENGERTAGRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDF TALIDFWDKV--- 580  
DB 1198 EAGHNETFSANAEDVSVTL-----DADLHFGVGYKEHRQITK-----DEANQG 1241  
  
QY 581 -----LEEAEAHLQSIPLPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLAN 636  
DB 1242 EGTGMHSGASEETIFNLGYDYCKOPTFTTVCVTEDKLLIQGA-YGFRAAKEESITLDAD 1300  
  
QY 637 AFFCTFPRRNAMKSEYSSYPDINFNRLFEGRSRKKPEKLKTLF-----CYFRRVTEKK 690  
DB 1301 LHFGVNYRDLLAMGLHAS-----NNEIEGTGMRSTASEETIFNLAYDYC-----KQ 1346  
  
QY 691 PTGLVTFTQSLSEDFPWECEKPLTRLHVTVYEGTIEENGQGMQLQVDFANRFVGGVTS 750  
DB 1347 PTEFRTVSVSEDHQFVHGAFGFRAVGEHHI--ETQVLE-----LQARMVEVMVEGVSVHL 1399  
  
QY 751 GLVQEEIRFLINPELLISRLFTFVLVDHNECLIIITGTQYSEYTGAEYVWS--RSHEDG 808

DB 1400 ARRHEDEPFV-----LYTEVIEET---IIRVDEQLEKKTTVIETEQAASEVQKREKG 1447  
  
QY 809 SERDDCERRCT-----EIVADIALHFRRYLDQFVP-----EKMRRE 844  
DB 1448 EERKBEKRVSFAAEVQEKIMEAIDKSLGLDTSMEVEPAFQKPSIHKPMKERRSRD 1507  
  
QY 845 LNKAYCGFLRPGVSSSENLSAVATG----- 868  
DB 1508 LRQNAAPAFKPVRRNSLLQALAGSPHNIPHFKTLDDIVKAIKHAGLEYSNLIIFGIDYTK 1567  
  
QY 869 -NWGCGAFGGDAR-----LKALIQI-----LAAAAAERDVVYFTFGDSELMRDI 911  
DB 1568 SNFYQGERTFDKRPLHTIDPAEMNPYQQVIOIVGKTLSSFDADGQIPAYGFGDEE----- 1622  
  
QY 912 YSMH-IF-LTERKLTVGDV--YKLLRYYNEECRNCSTPGPDIKLYPFIYHAVESCAE 965  
DB 1623 FTDHGIFNIAERYDLEKDCNGCFEVLRVYNEVTPTIEMSGP-TNFVPLIDRAIEICKE 1679  
  
RESULT 12  
S65236  
probable membrane protein YPL217c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein P1760  
C;Species: Saccharomyces cerevisiae  
C;Date: 10-Dec-1994 #sequence\_revision 31-May-1996 #text\_change 19-Apr-2002  
C;Accession: S65236  
R;Rieger, M.; Mueller-Auer, S.; Schaefer, M.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S65202  
A;Accession: S65236  
A;Molecule type: DNA  
A;Residues: 1-1183 <RIE>  
A;Cross-references: EMBL:Z73573; NID:G1370449; PID:e246934; PID:G1370450; GSPDB:GN00016;  
A;Experimental source: strain S288C (AB972)  
C;Genetics:  
A;Gene: SGD:BMS1; MIPS:YPL217C  
A;Cross-references: SGD:S0006138  
A;Map position: 16L  
C;Keywords: transmembrane protein  
F;126-142/Domain: transmembrane #status predicted <TMM>  
  
Query Match 3.1%; Score 159.5; DB 2; Length 1183;  
Best Local Similarity 18.3%; Pred. No. 0.1;  
Matches 156; Conservative 116; Mismatches 274; Indels 307; Gaps 36;  
  
QY 33 SRQRRVLDPKD-----AHVQFRVPPSPSPACVPQ-AGQHRGSAATSL 72  
DB 341 TRRRKRLLDDKDLIYAPMSDVGVLMKDQAVYIDIGKKNEEPSFVQGERGEKLMTGL 400  
  
QY 73 VFQKQTTTSMWDTKGI-----KTAESSELDSEKNNNNTRIE-----SMSSSVQKD 116  
DB 401 QSVEQSIAEKFDGVGLQFLFNGTELHEVADHEGMD-VESGEESIEDDEGKSKGRTSLRKP 459  
  
QY 117 NFYQHNV-EKLVNVSQLSLDKSLTEKSTQYLNQHQTAAAMCKWQNEGKHTEQLLESEPQT 175  
DB 460 RIYGKPVQBEDADIDNLPDDEE-----PYTND-----DDVQDSEPRMV 497  
  
QY 176 TLVPEQFSNANIDRSPQNDHSDTDSEENRDNQFLTTVKLANAKQTTEDEHA----- 228  
DB 498 EI---DFNNTGEQGAELALETDFSEFESEDEFSWERTA--ANKLKTESKRTWNIGKL 552  
  
QY 229 --REAKSHQCKSKSCHPGEDCASCQQDEIDVVPKSPLSVDVSGEDV-----GTGSKN 277  
DB 553 IYMDNISPEECIRRRWR-GEDDDS--KDESDI-----EEDVDDDFRKKDGTVTKE 599  
  
QY 278 DNK--LIHQESCLGNSPPFEKESEPEPMDV-----DNKNSQDSEADEETS 323  
DB 600 GNKDHAVDLEKFPVYFDTFEKLAKKWKSVDAIKERFLGAGILGNDNKTSDSNEGGEELY 659  
  
QY 324 PGFDEQEDGSSSQTANKPSRFQARDAD-----IEFRKRY 357  
DB 660 GDFELEDGPNPSEQAEDNSDKESEDEDEDNEDTNGDDDNSTFNFDAAEKKDLTMEQEREMN 719



```
QY 358 STKGGEVRLHFQFEGESRTGMNDLN-----AKLPGNISLNVEC-----RNS 400
Db 720 AAKKEKLAQAEIEEGENFKEDDENNEYDTWYELQAKISKQLEINNIEYQEMTPEQRQR 779
QY 401 KQHGGKDSKITDHLMLRP-----TKHQRTERKIPKYVPPHLSDDKKWLQ---KAEDRRKEQWE--- 429
Db 780 IEGFKAGSYVRIIVFEKVPMEFVKNFNPKFPIVMGGLLPTEIKFIVKARLRR-HRWHHKKI 838
QY 430 -----TKHQRTERKIPKYVPPHLSDDKKWLQ---TPIEEMRR 463
Db 839 LKTNDDLVLGLWRRFQTLPIYTTTDSRTRMLKTYTPEHTYCNAAFYGPLCSP-----N 893
QY 464 MPRCGIRLPLLRPSAN-----HTVTIRVDLLRAGEVPKPPFTHYKOLWMDNKHVK- 512
Db 894 TPFCGVQIVANSDTGNFRIATGIVEIDVNIIEIVKKLKL-VGFP--YKIFKNTAFIKD 950
QY 513 MPCSEONLYPVEDENGERTAGSRWELIQTALLNKFTRPQN-----LKDAILKYNVAYSKK 567
Db 951 MFSSAMEVARFEGAQIKTVSGIRGE-IKRAL-----SKPEGHYRAAFEDKILMSDIVILRS 1005
QY 568 WFTALIDFWDKVLBEAEQAHLQSYILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQ 627
Db 1006 WYFVRVKFPYNPVT-----LLKKEKTEWKGLRLT 1035
QY 628 EQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRL-----FEGRS----- 669
Db 1036 GQI-----RAAMNLETPSNPDSAYHKIERVERHFENGLKVPKAVQKELPFK 1080
QY 670 ----SRKPEKIKT 678
Db 1081 SQIHQMKPQKKKT 1093

RESULT 13
S38170
SRP40 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR092c; protein YKR412a
C:Species: Saccharomyces cerevisiae
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C:Accession: S38170; S40645; S37702
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38170
A:Molecule type: DNA
A:Residues: 1-406 <BAL>
A:Cross-references: EMBL:Z28317; NID:G486580; PID:G486581; MIPS:YKR092c
A:Experimental source: strain S288C
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Ji
Yeast 9, 1349-1354, 1993
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo
A:Reference number: S40644; MUID:94205265; PMID:8154186
A:Accession: S40645
A:Molecule type: DNA
A:Residues: 1-406 <BOU>
A:Cross-references: EMBL:X73541; NID:G450550; PIDN:CAA51946.1; PID:G450552
A:Experimental source: strain S288C
R:Lalo, D.; Carles, C.; Sentenac, A.; Thuriaux, P.
submitted to the EMBL Data Library, May 1993
A:Description: Interactions between three common subunits of yeast RNA polymerases I and
A:Reference number: S37702
A:Accession: S37702
A:Molecule type: DNA
A:Residues: 1-399, 'N', 401-406 <LAL>
A:Cross-references: EMBL:L11275; NID:G295670; PID:G295671
C:Genetics:
A:Gene: SGD:SRP40
A:Cross-references: SGD:S0001800; MIPS:YKR092c
A:Map position: 11R

Query Match 3.1%; Score 159; DB 2; Length 406;
Best Local Similarity 17.7%; Pred. No. 0.025;
```

```
Matches 61; Conservative 80; Mismatches 179; Indels 24; Gaps 4;

QY 39 LDPKDAHVQFRVPPSSPACVPGQAGQHRGSATSLVFKQKTTITSMWDTKGITAESESLDS 98
Db 14 LSVKEKEIEEKSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 73
QY 99 KENNTRIESMMSSVQKDNFYQHNVEKLVNVSLDKSLTEKSTQYLNQHTAAAMCK-W 157
Db 74 ESSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSESEDETKKRAES 133
QY 158 QNEGKHTEQLLESEPQ-----TVTLVPEQFSNANIDRSQPNDHSDTDSEENR 205
Db 134 DNEDAKETKAKTEPESSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 193
QY 206 DNQQLFTTVKLANAKQTTEDEHAREAKSHQKSKSCHPGEDECASCQQQDEIDVVPKPLSD 265
Db 194 DSQS-----SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSDS 246
QY 266 VGSSEVDVGTGSKNDNKLIRQESCLGNSPPPEKESEPESPMDVDNSKNSCQDSEADEETSPG 325
Db 247 SGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 302
QY 326 FDEQEDGSSSSQTANKPSRFQARDADIEFRKRYSTKGGEVRLHFQ 369
Db 303 NESTPSASSSSSANKLNIPAGTDEIKEGQKHFSRVDRSKINF 346

RESULT 14
A47328
natural killer cell tumor-recognition protein - human
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumc
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C:Accession: A47328
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cells.
A:Reference number: A47328; MUID:93133824; PMID:8421688
A:Accession: A47328
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1403 <AND>
A:Cross-references: GB:L04288; NID:gl81251; PIDN:AAA35734.1; PID:gl81252
A:Experimental source: NK killer cells from adult blood
A>Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIIP:122800)
C:Genetics:
A:Gene: GDB:NKTR
A:Cross-references: GDB:137171; OMIM:161565
A:Map position: 3p23-3p21
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>

Query Match 3.1%; Score 159; DB 1; Length 1403;
Best Local Similarity 19.6%; Pred. No. 0.14;
Matches 123; Conservative 93; Mismatches 225; Indels 188; Gaps 23;

QY 76 QKTITSMWDTKGITAESESLDSK-----ENNTRIESMMSSVQKD---NIFYQHNVEKLVN 128
Db 729 QSTYKYSYDRSESSSPRRSRSSRSYRSYTRSRSLASSHSRSPSSSRSHSRNKYSYD 788
QY 129 VSQSLDKSLTEKSTQYLNQHTAAAMCKQWNEKGKHTQLLESEPQTVTLVPEQFSNAN-- 186
Db 789 HSQCSRSSSYTSISS-----DDGRRARRLRSSGKNSVSHKHHSSSEK 833
QY 187 -----IDRSPQNDHSDTDSEENRDNQQLFTTVKLANAKQTTEDEHAREAKSHQK 237
Db 834 TLHSKYVKGRDRSSCVKYSERSSSLDYSSDSEQSSVQATQSAQEKQCGME-RTHNKQ 892
QY 238 SKSCHPGEDECASCQQQDEIDVVPKPLSDVGSSEVDVGTGSKNDNKLIRQESCLGNSPPPEKE 297
Db 893 EK--NRGEE-KSKSERECPHSKKRTTLKENLSDHLRNGSKPKRK-----NYAGSKWD 940

Query Match 3.1%; Score 159; DB 2; Length 406;
Best Local Similarity 17.7%; Pred. No. 0.025;
```



QY 298 SEPESPMDVNSKNSQDSEADEETSPGFDEQEDGSSQTANKPSRFQARDADIEFRKRY 357  
D 941 SESNSERDV--TKNSKNDSH-----PSSDKKEGEATSDSESEVS-----EIHKVKP 985  
QY 358 STKGGEVRLHFQEGGESRTGMNDLN-----AKLPGNISLNVCECRNSK-----QH 403  
D 986 TTK-----SSTNTSLPDDNGAMKSSKQRTSTSDSEGSCSNSENNRGKPKQKH 1033  
QY 404 GKDKSKITDHLMLRPLKAEDRRKEQWETKHQTERKIPKYVPPHLSPDKKWLGTPLEEMRR 463  
D 1034 GSKENLKREH---TKKVKEKLGGKKKKHKAAPKRAQAFHWQP-----1072  
QY 464 MPRCGLRPLLRPSANHTVIRVDLLRAGEVPPKPPPTHYKDLWDNKHVKMPCSEQNLYPV 523  
D 1073 -----PLFGESEEEEDDKQVTQESKEKV---1098  
QY 524 EDENGERTAGRWELIQTALLNKFTRPQNLKDALIKYNVA---YSKKWDTALIDFWDK 579  
D 1099 -SENNE-----TIKDNILKTEKSSEEDLSGKHDTVTVSSDLQ 1135  
QY 580 VLEEAEQAHLYSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAF 639  
D 1136 FTKDDSKLSISPTALNTEENVA-CLQNI-----QHVEESVPNGVEDVLQTDNMEI 1185  
QY 640 CTFPRRNAMKSEYSS-----YPDIN 660  
D 1186 CT-PDRSSPAKVEETSPLGNARLDTPDIN 1213  
RESULT 15  
SJHUA  
C;Species: Homo sapiens (man)  
C;Date: 17-May-1985 #sequence revision 30-Jun-1992 #text change 16-Jul-1999  
C;Accession: A35716; A45755; A23533; A91528; A42872; A02965; S13138; A05282; A23965; A27  
R;Sahr, K.E.; Laurila, P.; Kotula, L.; Scarpa, A.L.; Coupal, E.; Leto, T.L.; Linnenbach,  
J. Biol. Chem. 265, 4434-4443, 1990  
A;Title: The complete cDNA and polypeptide sequences of human erythroid alpha-spectrin.  
A;Reference number: A35716; MUID:90170949; PMID:1689726  
A;Accession: A35716  
A;Molecule type: mRNA  
A;Residues: 1-2429 <SAH>  
A;Cross-references: GB:J05244  
R;Sahr, K.E.; Tobe, T.; Scarpa, A.; Laughinghouse, K.; Marchesi, S.L.; Agre, P.; Linnenb  
J. Clin. Invest. 84, 1243-1252, 1989  
A;Title: Sequence and exon-intron organization of the DNA encoding the alpai domain of  
A;Reference number: A45755; MUID:9009318; PMID:2794061  
A;Accession: A45755  
A;Molecule type: DNA  
A;Residues: 1-394,'G',396-533 <SA2>  
A;Cross-references: GB:M29983  
A;Note: the authors translated the codon GGT for residue 395 as Ala  
R;Linnenbach, A.J.; Speicher, D.W.; Marchesi, V.T.; Forget, B.G.  
Proc. Natl. Acad. Sci. U.S.A. 83, 2397-2401, 1986  
A;Title: Cloning of a portion of the chromosomal gene for human erythrocyte alpha-spectr  
A;Reference number: A23533; MUID:86205962; PMID:3458204  
A;Accession: A23533  
A;Molecule type: DNA  
A;Residues: 320-450 <LIN>  
A;Cross-references: GB:M13233; NID:g182242; PIDN:AA83103.1; PID:g182243  
R;Curtis, P.J.; Palumbo, A.; Ming, J.; Fraser, P.; Clloe, L.; Meo, P.; Shane, S.; Rovera,  
Gene 36, 357-362, 1985  
A;Title: Sequence comparison of human and murine erythrocyte alpha-spectrin cDNA.  
A;Reference number: A91528; MUID:86083178; PMID:3000887  
A;Accession: A91528  
A;Molecule type: mRNA  
A;Residues: 1451-1687 <CUR>  
A;Cross-references: GB:M11049; NID:g338310; PIDN:AAA60569.1; PID:g553648  
R;Speicher, D.W.; Weglarz, L.; DeSilva, T.M.  
J. Biol. Chem. 267, 14775-14782, 1992  
A;Title: Properties of human red cell spectrin heterodimer (side-to-side) assembly and i  
A;Reference number: A42872; MUID:92340516; PMID:1634521  
A;Accession: A42872

A;Molecule type: protein  
A;Residues: 7-16;46-55;680-689;1047-1056;1921-1930 <SPI>  
R;Speicher, D.W.; Davis, G.; Marchesi, V.T.  
J. Biol. Chem. 258, 14938-14947, 1983  
A;Title: Structure of human erythrocyte spectrin. II. The sequence of the alpha-I domain  
A;Reference number: A92408; MUID:84087888; PMID:6654896  
A;Accession: A02965  
A;Molecule type: protein  
A;Residues: 7-601 <SPC>  
R;Speicher, D.W.; Davis, G.; Yurchenco, P.D.; Marchesi, V.T.  
J. Biol. Chem. 258, 14931-14937, 1983  
A;Title: Structure of human erythrocyte spectrin. I. Isolation of the alpha-I domain and  
A;Reference number: S13138; MUID:84087887; PMID:6654895  
A;Accession: S13138  
A;Molecule type: protein  
A;Residues: 7-92,'X',94-96,110-151;317-342;345-366,'Z',368-370,'X',372-373;387-434;452-4  
R;Lusitani, D.M.; Qtaishat, N.; LaBrake, C.C.; Yu, R.N.; Davis, J.; Kelley, M.R.; Fung,  
J. Biol. Chem. 269, 25955-25958, 1994  
A;Title: The first human alpha-spectrin structural domain begins with serine.  
A;Reference number: A38928; MUID:95014412; PMID:7929303  
A;Contents: annotation; determination of structural domain  
C;Comment: Spectrin associates with band 4.1 and actin to form the cytoskeletal superstr  
s of approximately 106 residues each.  
C;Genetics:  
A;Gene: GDB:SPTA1  
A;Cross-references: GDB:119601; OMIM:182860  
A;Map position: 1q21-1q21  
C;Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectrin/  
C;Keywords: actin binding; cytoskeleton; duplication; EF hand; erythrocyte; heterodimer;  
F;52-156/Domain: spectrin/dystrophin repeat homology <SP1>  
F;157-262/Domain: spectrin/dystrophin repeat homology <SP2>  
F;263-368/Domain: spectrin/dystrophin repeat homology <SP3>  
F;369-474/Domain: spectrin/dystrophin repeat homology <SP4>  
F;475-580/Domain: spectrin/dystrophin repeat homology <SP5>  
F;581-685/Domain: spectrin/dystrophin repeat homology <SP6>  
F;686-791/Domain: spectrin/dystrophin repeat homology <SP7>  
F;792-897/Domain: spectrin/dystrophin repeat homology <SP8>  
F;898-983/Domain: spectrin/dystrophin repeat homology <SP9>  
F;984-1031/Domain: SH3 homology <SH3>  
F;1081-1181/Domain: spectrin/dystrophin repeat homology <SP10>  
F;1182-1287/Domain: spectrin/dystrophin repeat homology <SP11>  
F;1288-1393/Domain: spectrin/dystrophin repeat homology <SP12>  
F;1394-1498/Domain: spectrin/dystrophin repeat homology <SP13>  
F;1499-1604/Domain: spectrin/dystrophin repeat homology <SP14>  
F;1605-1710/Domain: spectrin/dystrophin repeat homology <SP15>  
F;1711-1816/Domain: spectrin/dystrophin repeat homology <SP16>  
F;1817-1925/Domain: spectrin/dystrophin repeat homology <SP17>  
F;1926-2032/Domain: spectrin/dystrophin repeat homology <SP18>  
F;2041-2146/Domain: spectrin/dystrophin repeat homology <SP19>  
F;2155-2257/Domain: spectrin/dystrophin repeat homology <SP20>  
F;2270-2302/Domain: calmodulin repeat homology <EF1>  
F;2313-2345/Domain: calmodulin repeat homology <EF2>

Query Match 3.1%; Score 159; DB 1; Length 2429;  
Best Local Similarity 21.0%; Pred. No. 0.31;  
Matches 131; Conservative 99; Mismatches 216; Indels 178; Gaps 33;

QY 79 ITSWMDTKGIKTAESLSDSKENNNTRIESMSSVQKDNFYQHNVKLVNVSQI-SLDKS 137  
D 594 LKNWINKKK-KLADDE--DYKD-----IQNLSRVQKQVFEK--ELAVNKTQLENIQKT 643  
QY 138 LTE--KSTQYLNQHQTAAAMCK-----WQNEGKHTEQLLESEPTVTLVPE-----QFSNAN 186  
D 644 QQEMIEGGHYASDNVTTRLSEVASLW-----BELLEATKQKGTQLHEANQQLQFEN-- 694  
QY 187 IDRSFQNDHSDTSDSENRDNQFLTTTVKLANAKQTTEDEHAREAKSHQKCKSCHPGE 246  
D 695 -----NAEDLQRWLEDVEW-----QVTSEDYKGLAEVQNRLRKHGLLES 734  
QY 247 CASCCQDEIDVVPKSPISDVGS--EDVGTGSKNDKNKILRQESCLGNSPPFEKESEP---- 300  
D 735 AVAARQDQVDI-----LTDLAAYFEIIGHFDPDSKDIR-ARQESLVCVCR---FEALKEPLATR 785

Search completed: May 26, 2004, 18:47:29  
Job time : 31.8489 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 15.6372 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-4  
Perfect score: 5190  
Sequence: 1 MNAGPGCEPCTKATRWGAAT.....YHAVESCAETADHSGQRTGT 976

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description        |
|------------|-------|-------------|--------|--------------|--------------------|
| 1          | 178   | 3.4         | 1359   | 1 ATRX CAEEL | Q9u7e0 caenorhabdi |
| 2          | 174   | 3.4         | 5147   | 1 PCLO HUMAN | Q9y6v0 homo sapien |
| 3          | 170.5 | 3.3         | 5038   | 1 PCLO_MOUSE | Q9qyx7 mus musculu |
| 4          | 168   | 3.2         | 5085   | 1 PCLO_RAT   | Q9jks6 rattus norv |
| 5          | 162   | 3.1         | 2492   | 1 ATRX PANTR | Q7ygm4 pan troglod |
| 6          | 161   | 3.1         | 2476   | 1 ATRX_MOUSE | Q61687 mus musculu |
| 7          | 161   | 3.1         | 2492   | 1 ATRX_HUMAN | P46100 homo sapien |
| 8          | 161   | 3.1         | 2492   | 1 ATRX_PONPY | Q7ygm3 pongo pygma |
| 9          | 159.5 | 3.1         | 1183   | 1 BMS1 YEAST | Q08965 saccharomyc |
| 10         | 159   | 3.1         | 406    | 1 SR40 YEAST | P32583 saccharomyc |
| 11         | 159   | 3.1         | 1462   | 1 NKCR_HUMAN | P30414 homo sapien |
| 12         | 159   | 3.1         | 2418   | 1 SPCA_HUMAN | P02549 homo sapien |
| 13         | 157.5 | 3.0         | 1311   | 1 ATRX_DROME | Q9gqn5 drosophila  |
| 14         | 155   | 3.0         | 1138   | 1 BM2K_MOUSE | Q91296 mus musculu |
| 15         | 154.5 | 3.0         | 646    | 1 SG1_BOVIN  | P23389 bos taurus  |
| 16         | 154.5 | 3.0         | 3259   | 1 GOB1_HUMAN | Q14789 homo sapien |
| 17         | 154   | 3.0         | 1972   | 1 MYHB_MOUSE | Q08638 mus musculu |
| 18         | 153.5 | 3.0         | 1878   | 1 BRC1_CANFA | Q95153 canis famil |
| 19         | 153   | 2.9         | 1031   | 1 YDG9_SCHPO | Q10496 schizosacch |
| 20         | 153   | 2.9         | 6632   | 1 UN89_CAEEL | O01761 caenorhabdi |
| 21         | 152   | 2.9         | 1447   | 1 BUD4 YEAST | P47136 saccharomyc |
| 22         | 151   | 2.9         | 2415   | 1 SPCA_MOUSE | P08032 mus musculu |
| 23         | 150.5 | 2.9         | 1312   | 1 RA50 YEAST | P12753 saccharomyc |
| 24         | 150.5 | 2.9         | 1957   | 1 SPOF_SCHPO | Q10411 schizosacch |
| 25         | 150.5 | 2.9         | 1972   | 1 MYHB_RABIT | P35748 oryctolagus |
| 26         | 150   | 2.9         | 2137   | 1 SPCB_HUMAN | P11277 homo sapien |
| 27         | 150   | 2.9         | 3969   | 1 HRX_HUMAN  | Q03164 homo sapien |
| 28         | 149.5 | 2.9         | 1210   | 1 AP4_HUMAN  | P51825 homo sapien |
| 29         | 148.5 | 2.9         | 472    | 1 YWIE_CAEEL | Q23525 caenorhabdi |
| 30         | 148.5 | 2.9         | 677    | 1 SG1_HUMAN  | P05060 homo sapien |
| 31         | 148.5 | 2.9         | 1928   | 1 MYS1 YEAST | P08964 saccharomyc |
| 32         | 148   | 2.9         | 644    | 1 NFM_RABIT  | P54938 oryctolagus |
| 33         | 148   | 2.9         | 1679   | 1 GCC2_MOUSE | Q8chg3 mus musculu |

|    |       |     |      |              |                    |
|----|-------|-----|------|--------------|--------------------|
| 34 | 146.5 | 2.8 | 667  | 1 CYL1_BOVIN | P35662 bos taurus  |
| 35 | 146.5 | 2.8 | 3214 | 1 BPAL_HUMAN | Q03001 homo sapien |
| 36 | 146   | 2.8 | 5120 | 1 PCLO_CHICK | Q9pu36 gallus gall |
| 37 | 144.5 | 2.8 | 1980 | 1 MY9B_RAT   | Q63358 rattus norv |
| 38 | 144.5 | 2.8 | 2319 | 1 AKA6_HUMAN | Q13023 homo sapien |
| 39 | 144   | 2.8 | 1637 | 1 MRSP_STAAU | P80544 staphylococ |
| 40 | 143.5 | 2.8 | 743  | 1 ABRA_PLAFC | P22620 plasmodium  |
| 41 | 143.5 | 2.8 | 1338 | 1 ACIN_MOUSE | Q9jix8 mus musculu |
| 42 | 143.5 | 2.8 | 1969 | 1 MYS1_CAEEL | P12844 caenorhabdi |
| 43 | 143.5 | 2.8 | 4835 | 1 MDN1_GTALA | Q8t5t1 giardia lam |
| 44 | 143   | 2.8 | 1085 | 1 YAF1_SCHPO | Q09863 schizosacch |
| 45 | 143   | 2.8 | 1433 | 1 REST_CHICK | O42184 gallus gall |

ALIGNMENTS

RESULT 1

ATRX CAEEL  
ID ATRX CAEEL STANDARD; PRT; 1359 AA.  
AC Q9U7E0; O02061;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).  
GN XNP-1 OR B0041.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=99365296; PubMed=10433961;  
RA Villard L., Fontes M., Ewbank J.J.;  
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to  
the human XNP/ATR-X gene."  
RL Gene 236:13-19(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Fulton R., Wohldmann P.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies  
gene expression by affecting chromatin (Potential).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
EMBL; AF134186; AAD55361.1; --  
EMBL; AF000196; AAC24256.1; --  
PIR; T34036; T34036.  
WormPep; B0041.7; CE17314.  
InterPro; IPR001410; DEAD.  
InterPro; IPR001650; Helicase\_C.  
InterPro; IPR000330; SNF2\_N.  
Pfam; PF00271; helicase\_C; 1.  
Pfam; PF00176; SNF2\_N; 1.  
SMART; SM00487; DEXDC; 1.  
SMART; SM00490; HELIC; 1.  
PROSITE; PS00690; DEAD ATP HELICASE; FALSE NEG.  
DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;  
DNA-binding.  
NP\_BIND 496 503  
SITE 636 639 DEAH BOX.  
DOMAIN 67 70 POLY-ASP.  
DOMAIN 266 272 POLY-GLU.

|                       |   |  |                 |                             |
|-----------------------|---|--|-----------------|-----------------------------|
| FT                    | DOMAIN  | 276  | 281             | POLY-LYS.                   |
| FT                    | DOMAIN  | 372  | 375             | POLY-LYS.                   |
| FT                    | DOMAIN  | 603  | 608             | POLY-LYS.                   |
| FT                    | DOMAIN  | 859  | 862             | POLY-LYS.                   |
| FT                    | CONFLICT  | 479  | 479             | C -> F (IN REF. 2).         |
| SQ                    | SEQUENCE  | 1359   | AA; 156191      | MW; EB4342547D4F4B64 CRC64; |
| Query Match           |   |  |                 |                             |
| Best Local Similarity |   | 3.4%;  | Score 178;      | DB 1; Length 1359;          |
| Matches 147;          |   | Conservative 118;  | Mismatches 279; | Indels 224; Gaps 33;        |
| QY                    | 86  | KGTKAESESLDSKNNNTRIESMMSSVQKDNFYQHNVKLVNVSQSLDKSLTEKSTQY     | 145             |                             |
| Db                    | 81  | KSRKRAKSES-ESDESDEEDRKSKSKKKVDQKKKBSKKKRTTSSEDESDEREQ-       | 138             |                             |
| QY                    | 146   | LNQHTAAMCKWQNEGKHTQELLESEPQTTLVPEQFSNANIDRSPQNDHSDTDSSENR    | 205             |                             |
| Db                    | 139   | -KSKKSKTKKQTSSESSE--ESEEKVKKSKKKNKESVKKRAETSEESDEDEKPSK      | 194             |                             |
| QY                    | 206   | DNQQLTTVKLANAKQTTEDEHAREAKSHQKCSKCHPGEDCASCQODEIDVVPKSPLSD   | 265             |                             |
| Db                    | 195   | KSKKGLK--KKAKESESESEDEKEVKKSKKKSKKVVKKSESEDEAPEKKKTEKRRSK    | 252             |                             |
| QY                    | 266   | VGSEDVGTGSKNDNKLIRQESCLGNSPPPEKESEP-----ESPMDVDSKNSCQSEADEE  | 321             |                             |
| Db                    | 253   | TSSESESESEKSDEE--EE-----EKESPPKPKKKPLAVKKLSSDESESDVE         | 300             |                             |
| QY                    | 322   | TSP-----GFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTKGGEVRLHF         | 368             |                             |
| Db                    | 301   | VLPQKKRGAVTLISDEDEKQSESEASDVEEKSVKKA-----KKQESSE-----        | 348             |                             |
| QY                    | 369   | QFEGGESRTGMNDLNKLPGNISSLVNVECRNSKQHG--KKDSKITDHLMLRPLKAEDRKE | 426             |                             |
| Db                    | 349   | --SGSDSESGSITVNRK-----SKKKEKPEKKKKGIIMDSKQLKETIDAERAERKRK    | 400             |                             |
| QY                    | 427   | QWETKHQR-----TERKIPKYVPPHLSPDKKWLGTPIEEMRMMPR                | 466             |                             |
| Db                    | 401   | RLEKKQKEFGIVLEEGEDLTEMLTGTSSQRLKSVV---LDPD-----SSTVDEESKXP-  | 452             |                             |
| QY                    | 467   | CGIRLPLLRPSANHTVTIRVDLLRAGEVPKPPFTHYKDLWKNKHVMKPCSEQNLYPVEDE | 526             |                             |
| Db                    | 453   | -----VEVHNSLVR---ILKPHQAH-----GIQFMYDCACE                    | 480             |                             |
| QY                    | 527   | NGER--TAGSRWELIQTALNKEFRPQNLKDAILKYNVAYSKKNDFTALIDFWDKVLLEA  | 584             |                             |
| Db                    | 481   | SILRLDTEGS-----GGILAHCMGLGKT---LQVITFLHTVL---                | 514             |                             |
| QY                    | 585   | EAQHLYQSILPDMVKIALCLP-NICTQPIPLLKQKNHSITMSQEQIASLLANAFCTFP   | 643             |                             |
| Db                    | 515   | ----MHEKIGEKCKRVLVVPKQNVINWF----KEFQKWLVDNDEELDTIDVN-----    | 559             |                             |
| QY                    | 644   | RNAKMKSEYSSYPDINFNR--LFEGRSSRKPEKLTFLCFVFRVVT-----EKKPTGLV   | 695             |                             |
| Db                    | 560   | -----ELDSYKTIEDRRRALKAWHSSKTPSVMIIGYDLFRILTVEDDPKCKKPKNRN    | 611             |                             |
| QY                    | 696   | TFTRQSLDEF-----PEWERCE-----KPLTRLHVTYEGTIEENG                | 730             |                             |
| Db                    | 612   | RRLEKAKEDFRKYLQNPDPDMVVCDEAHKLNDDSAKSKMVKILTKRRICTGTPLQNN    | 671             |                             |
| QY                    | 731   | -----QGML--QVDFANRFVGGVTSAGLVQE-----EIRFL                    | 760             |                             |
| Db                    | 672   | LMEYHGMVNFVKPGLLGTGKTEFANRFV--NIINRGRTKASPLEVSFM             | 717             |                             |
| RESULT 2              |   |  |                 |                             |
| PCLO HUMAN            |   |  |                 |                             |
| ID                    | PCLO HUMAN                                      | STANDARD;  | PRT;            | 5147 AA.                    |
| AC                    | Q9Y6V0; O43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9; |  |                 |                             |
| DT                    | 28-FEB-2003 (Rel. 41, Created)                  |  |                 |                             |
| DT                    | 28-FEB-2003 (Rel. 41, Last sequence update)     |  |                 |                             |
| DT                    | 10-OCT-2003 (Rel. 42, Last annotation update)   |  |                 |                             |
| DE                    | Piccolo protein (Aczonin) (Fragments).          |  |                 |                             |
| GN                    | PCLO OR ACZ OR KIAA0559.                        |  |                 |                             |

|    |   |
|----|---|
| OS | Homo sapiens (Human).   |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;     |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.            |
| OX | NCBI_TaxID=9606;  |
| RN | [1]   |
| RP | SEQUENCE OF 1-759 FROM N.A.   |
| RC | TISSUE=Brain;   |
| RX | MEDLINE=99439764; PubMed=10508862;                                    |
| RA | Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,      |
| RA | Killmann M.W.;  |
| RT | "Aczonin, a 550-kd putative scaffolding protein of presynaptic active |
| RT | zones, shares homology regions with rim and bassoon and binds         |
| RT | profilin.";   |
| RL | J. Cell Biol. 147:151-162(1999).                                      |
| RN | [2]   |
| RP | SEQUENCE OF 552-4404 FROM N.A.  |
| RA | Kraemer J., Wollam C., Wohldmann P., McGrane B.;                      |
| RL | Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.               |
| RN | [3]   |
| RP | SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).                          |
| RC | TISSUE=Brain;   |
| RX | MEDLINE=98290545; PubMed=9628581;                                     |
| RA | Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,         |
| RA | Nomura N., Ohara O.;  |
| RT | "Prediction of the coding sequences of unidentified human genes. IX.  |
| RT | The complete sequences of 100 new cDNA clones from brain which can    |
| RT | code for large proteins in vitro.";                                   |
| RL | DNA Res. 5:31-39(1998).   |
| RN | [4]   |
| RP | SEQUENCE OF 4405-4439 FROM N.A.                                       |
| RC | TISSUE=Placenta;  |
| RX | MEDLINE=22388257; PubMed=12477932;                                    |
| RA | Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,              |
| RA | Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,   |
| RA | Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,     |
| RA | Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,       |
| RA | Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,         |
| RA | Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., |
| RA | Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,    |
| RA | Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., |
| RA | Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,   |
| RA | Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., |
| RA | Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,         |
| RA | Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  |
| RA | Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,       |
| RA | Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,              |
| RA | Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,                  |
| RA | Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,        |
| RA | Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;                   |
| RT | "Generation and initial analysis of more than 15,000 full-length      |
| RT | human and mouse cDNA sequences.";                                     |
| RL | Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).                   |
| RN | [5]   |
| RP | SEQUENCE OF 4405-5147 FROM N.A.                                       |
| RA | Kalicki J., Elliott G.;   |
| RL | Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.               |
| CC | -!- FUNCTION: May act as a scaffolding protein involved in the        |
| CC | organization of synaptic active zones and in synaptic vesicle         |
| CC | trafficking (By similarity).  |
| CC | -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By       |
| CC | similarity).  |
| CC | -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of     |
| CC | synaptic junctions (By similarity).                                   |
| CC | -!- ALTERNATIVE PRODUCTS:   |
| CC | Event=Alternative splicing; Named isoforms=2;                         |
| CC | Comment=Additional isoforms seem to exist;                            |
| CC | Name=1;   |
| CC | Isoid=Q9Y6V0-1; Sequence=Displayed;                                   |
| CC | Name=2;   |
| CC | Isoid=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,          |
| CC | VSP_003926, VSP_003927;   |
| CC | Note=No experimental confirmation available;                          |
| CC | -!- DOMAIN: C2 domain 1 is involved in binding calcium and            |



phospholipids. Calcium binds with low affinity but with high specificity and induces a large conformational change.

-|- SIMILARITY: Contains 2 C2 domains.

-|- SIMILARITY: Contains 1 PDZ/DHR domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; Y19188; CAB60727.1; --

EMBL; AC004903; AAD20936.1; --

EMBL; AC004986; AAD21789.1; --

EMBL; AB011131; BAA25485.1; --

EMBL; BC001304; AAH01304.1; --

EMBL; AC004082; AAB97937.1; --

PIR; T00634; T00634.

HSSP; P04410; 1A25.

Genew; HGNC:13406; PCLO.

MIM; 604918; --

GO; GO:0005856; C:cytoskeleton; NAS.

GO; GO:0045202; C:synaptic junction; ISS.

GO; GO:0005509; F:calcium ion binding; ISS.

GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.

GO; GO:0005522; F:profilin binding; ISS.

GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.

GO; GO:0016080; P:synaptic vesicle targeting; ISS.

InterPro; IPR000008; C2.

InterPro; IPR001565; Synaptotagmin.

PRINTS; PR00360; C2DOMAIN.

PRINTS; PR00399; SYNAPTOTAGMIN.

SMART; SM00239; C2; 2.

PROSITE; PS00499; C2\_DOMAIN\_1; 1.

PROSITE; PS00004; C2\_DOMAIN\_2; 2.

Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger; Repeat; Alternative splicing.

NON\_TER 1 1

DOMAIN 400 465

10 X 10 AA TANDEM APPROXIMATE REPEATS OF

P-A-K-P-Q-P-Q-P-X.

C4-TYPE (POTENTIAL).

C4-TYPE (POTENTIAL).

POLY-PRO.

PDZ.

C2 DOMAIN 1.

C2 DOMAIN 2.

S -> SGNGLGIRIVGGKEIPGHSGEIGAYIAKILPGSSAE

QTGKLMG (in isoform 2).

/FTid=VSP 003923.

K -> KPTDGTKVWSHPITGEIQ (in isoform 2).

/FTid=VSP 003924.

G -> GQVMVQNAS (in isoform 2).

/FTid=VSP 003925.

TAHKS -> SKRRK (in isoform 2).

/FTid=VSP 003926.

Missing (in isoform 2).

/FTid=VSP 003927.

SEQUENCE 5147 AA; 563537 MW; CD5D84990498CD3C CRC64;

Query Match 3.4%; Score 174; DB 1; Length 5147;

Best Local Similarity 19.7%; Pred. No. 0.099;

Matches 140; Conservative 120; Mismatches 251; Indels 198; Gaps 32;

9 PCTKATRWGAATTSPAASDARSPSPQRRLVDPKDAHVQVRVPPSSPACVPGQAGQHRGS 68

1227 PCT-----AKPDQEKEDKSDTSSSQ-----PKSPQGLSDTGYSDDGI 1265

69 ATSLVFKQKITSWMDT-----KGIK--TAESSELDSENNNTRIESMSSVQKDNFYQ 120

1266 SSSL-----GEIPSLIPTDEKILGLKDKDSFSQESSPSSPSDLAKLESTVLSILE----- 1316

Qy 121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHTAAMCKWQNEKGKHTQLLESEPTVTLVPE 180

Db 1317 -----AQASTLADEKS--EKKTQ---PHEVSP--EQPKDQEKTKQSL--SETLEITISEE 1361

Qy 181 QFSNANIDR-----SPQNDHSDTDSEENRDNQQLTTVK--LANAKQTTDEHA--- 228

Db 1362 EIKESQERKDTFKKDSQQDIPSSKDHEKSEFVDDITTRREPYDSVESESESPVPQ 1421

Qy 229 REAKSHQCKSCHPGEDCASCQDEIDVVPKSPLSDVSESDVGTGSKNDKLRQESCL 288

Db 1422 RKRTSVGSSSSDEYKQSDSQSGSEEDFIRKQ--IIEMSADEDASGEDDEFIRNQLKEI 1480

Qy 289 GNSPPFEKESEPE-----SPMDVNSKNSCQDSEADEETSFGFDEQE 330

Db 1481 SSSTESQKKEETKGKITAGKHRLTRKSTSIDEDAGRHSWHDEDE---AFDESP 1536

Qy 331 DGSSSQATANKPS-----RFQARDADIEFRKRYSTKGGEVRLHFQF----- 370

Db 1537 ELKYRETKSQESEELVVTGGGLRRFKTIELNSTIADKYSAESSQKTSLYFDEEPELEM 1596

Qy 371 ----EGGESRT---GMNDLNAK-----LPGNISSLNVECRNSKQHGKDKSK-----IT 411

Db 1597 ESLTDSPEDRSRGEGSSSLHASSFTPGTSPTSVSSLDSDSDSSPHKKGESQQRKARHR 1656

Qy 412 DHLMLPKAEDR-----RKEQWETKHQTERKI----- 439

Db 1657 PHGPLLTPTIEDSSEEEELREELKEQEKQREIEQQQRKSSSKSKDKDELRAQRRE 1716

Qy 440 -PKYVPPHLSPDKKWLGTPIEMRRMPCGIRLPILRPS--ANHTVTIRVDLLRAGEVPKP 497

Db 1717 RPKTPPSNLSPIED--ASPEELRQAAEME---ELHRSSCSEYSPSIESD-----PEG 1764

Qy 498 FPTHYKDLWD-NKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLLKDA 556

Db 1765 FEISPEKIIIEVKYKLP-TAVSLYSPTDEQ-----SIMQEGSQKALKSA 1809

Qy 557 ILKYNVAYSKKWDFTL--IDFWDKVLSEAE-----AQHLYQSILPD 596

Db 1810 EEMYEMMHKTHKYKAFPAANERDEVEFEKEPLYGGMLIEDYIYESLVED 1858

RESULT 3

PCLO MOUSE

ID PCLO MOUSE STANDARD; PRT; 5038 AA.

AC Q9QYX7; Q9QYX6; Q9QZJ0;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-derived HLMN protein).

DE PCLO OR ACZ.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING, TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.

RC TISSUE=Brain;

RX MEDLINE=99439764; PubMed=10508862;

RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E., Killmann M.W.;

RA "Aczonin, a 550-kd putative scaffolding protein of presynaptic active zones, shares homology regions with rim and bassoon and binds profilin."

RL J. Cell Biol. 147:151-162(1999).

RN [2]

RP REVISIONS.

RA Killmann M.W.;

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 4502-4682 FROM N.A.

RC TISSUE=Brain;  
RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.  
RN [4]  
RP INTERACTION WITH RIMS2.  
RX MEDLINE=22384373; PubMed=12401793;  
RA Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,  
RT Sasaki T., Tajima N., Iwanaga T., Seino S.;  
RT "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of  
RT CAMP-GEFII.Rim2.Piccolo complex in CAMP-dependent exocytosis.";  
RL J. Biol. Chem. 277:50497-50502(2002).  
CC -!- FUNCTION: May act as a scaffolding protein involved in the  
CC organization of synaptic active zones and in synaptic vesicle  
CC trafficking.  
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin.  
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of  
CC synaptic junctions.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q9QYX7-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9QYX7-2; Sequence=VSP\_003928, VSP\_003929;  
CC TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in  
CC stomach. Not detected in other tissues analyzed including adrenal  
CC gland, testis and pancreas.  
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and  
CC phospholipids. Calcium binds with low affinity but with high  
CC specificity and induces a large conformational change.  
CC -!- SIMILARITY: Contains 2 C2 domains.  
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; Y19185; CAB60731.2; -;  
EMBL; Y19186; CAB60732.2; -;  
EMBL; AF181269; AAD55786.2; -;  
HSSP; P04410; 1A25.  
MGD; MGI:1349390; Pclo.  
GO; GO:0045202; C:synaptic junction; IDA.  
GO; GO:0005509; F:calcium ion binding; ISS.  
GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.  
GO; GO:0005522; F:profilin binding; IDA.  
GO; GO:0019933; P:cAMP-mediated signaling; IDA.  
GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.  
GO; GO:0030073; P:insulin secretion; IDA.  
GO; GO:0017157; P:regulation of exocytosis; IDA.  
GO; GO:0016080; P:synaptic vesicle targeting; NAS.  
InterPro; IPR000008; C2.  
InterPro; IPR001478; PDZ.  
InterPro; IPR008899; Znf\_piccolo.  
Pfam; PF00168; C2; 2.  
Pfam; PF00595; PDZ; 1.  
Pfam; PF05715; Zf\_piccolo; 2.  
SMART; SM00239; C2; 2.  
SMART; SM00228; PDZ; 1.  
PROSITE; PS00499; C2\_DOMAIN\_1; 1.  
PROSITE; PS50004; C2\_DOMAIN\_2; 2.  
PROSITE; PS50106; PDZ; 1.  
Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;  
Repeat; Alternative splicing.  
DOMAIN 371 470 10 X 10 AA TANDEM APPROXIMATE REPEATS OF  
P-A-K-P-Q-P-Q-P-X.  
C4-TYPE (POTENTIAL).  
C4-TYPE (POTENTIAL).  
POLY-PRO.  
PDZ.  
ZN\_FING 502 526  
ZN\_FING 967 990  
DOMAIN 2305 2329  
DOMAIN 4394 4488

|   |   |  |            |                                |
|---|---|--|------------|--------------------------------|
| FT  | DOMAIN  | 4607   | 4705       | C2 DOMAIN 1.                   |
| FT  | DOMAIN  | 4922   | 5012       | C2 DOMAIN 2.                   |
| FT  | VARSPPLIC   | 4829   | 4833       | TKPTN -> SKRRK (in isoform 2). |
| FT  | VARSPPLIC   | 4834   | 5038       | /FTId=VSP_003928.              |
| FT  | VARSPPLIC   | 4834   | 5038       | Missing (in isoform 2).        |
| FT  | SEQUENCE  | 5038   | AA; 547600 | MW; DADA460CF3B40888 CRC64;    |
| SQ  | SEQUENCE  | 5038   | AA; 547600 | MW; DADA460CF3B40888 CRC64;    |
| Query Match   |   |  |            |                                |
| Best Local Similarity 3.3%; Score 170.5; DB 1; Length 5038;         |   |  |            |                                |
| Matches 141; Conservative 122; Mismatches 271; Indels 187; Gaps 33; |   |  |            |                                |
| QY  | 5   | PGCEPCT-----KATRWGAATTSP-----AASDARSFPQRVLPDPKDAHVFQFRVPPSS  | 54         |                                |
| Db  | 1206  | PGATPQTLPKDRQKESR---DVTQPAEGTAKGREGPSKDRTEKEEDKSDTSSSQPKS    | 1262       |                                |
| QY  | 55  | PACVPGQAGQHRGSATSLVFKQKTTITSMMD---TKGIK--TAESSELDSENNNTRIES  | 108        |                                |
| Db  | 1263  | PQGLSDTGYSSDGLSGSLGEIPSLIPS--DEKDLLKGLKKDSFSQESSPSPSDLAKLES  | 1320       |                                |
| QY  | 109   | MMSSV---QKDNFYQHNVVEKLVNVSQSLDKSLTEKSTQYLNQHOHTAAMCKWQNEGKHT | 165        |                                |
| Db  | 1321  | TVLSILEAQASTLVGEKAEKKTQPKVSPQEQDQKQKTQTPSETRDISIE-----E      | 1372       |                                |
| QY  | 166   | QLLESEPQTTLVPEQFSNANIDRSPONDHSDTDSENRDNQOFLTITVKLANA-----K   | 220        |                                |
| Db  | 1373  | EIKESQEKVT-----SKDASAQGFPSRKEHKNPELVDDLSPRRASYDSVE           | 1419       |                                |
| QY  | 221   | QTTEDEHAREAKSHQ-----CSKSGHPGEDCASCQQDEIDVVPKPSLSDVSGEDVGTGS  | 275        |                                |
| Db  | 1420  | DSSESENSPVARRKRTSIGSSSEYKQEDSGSGEDE-DFIRKQ-IIEMSADEDASGS     | 1477       |                                |
| QY  | 276   | KNDNKLIRQESCLGNSPPFEKESEP-----ESPM-----VDNSKNSCQDS-----EA    | 318        |                                |
| Db  | 1478  | EDEEFIRSQLKEIGGVTESQKREETKKGKSPAGKHRLTRKSTSTFDDAGRRHSHWDE    | 1537       |                                |
| QY  | 319   | DEETSPGFDEQEDSSSQANKPS-----RFOARDADIEFRKRYSTKGGEVRL          | 366        |                                |
| Db  | 1538  | DDET---FDESPELKFRETQSQSEELVAVGGGLRRFKTIELNSTVTDKYAESQKKT     | 1594       |                                |
| QY  | 367   | HFQF-----EGGESRT---GMNDLNAK-----LPGNISSLNVECNKQHGK           | 405        |                                |
| Db  | 1595  | TLYFDEEPELEMESLTDSPEDRSRGEGSSSLHASSFTPGTSTSVSSLDEDSSSPSHKK   | 1654       |                                |
| QY  | 406   | KDSK-----ITDHLMLPKAEDR-----RKEQWETKHQTERKI-----              | 439        |                                |
| Db  | 1655  | GESKQQRKARHRSRGPLPTIEDSSEELREEELKEQEKQRELEQQQRKSSSKSKK       | 1714       |                                |
| QY  | 440   | -----PKYVPPHLSPKKWLGTPIEEMRMRPCGIRLPLLRPS-ANHTVTIR           | 485        |                                |
| Db  | 1715  | DKDELRAQRRRRPKTPPSNLSPIED--ASPTEELRQAAEME---ELHRSSCSEYSPSIE  | 1769       |                                |
| QY  | 486   | VDLLRAGEVPKPPPTHYKDLWD-NKHVMPCSEQNLYPVEDENGERTAGSRWELIQTALL  | 544        |                                |
| Db  | 1770  | SD-----PEGFEISPEKIEVKVYKLP-TAVSLYSPTDEQ-----SVM              | 1807       |                                |
| QY  | 545   | NKTRPQNLKDAILKYNVAYSKKWDFAL--IDFWDKVLSEAE-----AQHLYQSILP     | 595        |                                |
| Db  | 1808  | QKEGAQKALKSAEEMEYEMMHKPKYKAFPAANERDEVEFEKPLYGGMIEDIYVESLVE   | 1867       |                                |
| QY  | 596   | D 596  |            |                                |
| Db  | 1868  | D 1868   |            |                                |
| RESULT 4  |   |  |            |                                |
| PCLO RAT  |   |  |            |                                |
| ID  | PCLO RAT  | STANDARD;  | PRT;       | 5085 AA.                       |
| AC  | Q9JKS6; Q9JLT1;   |  |            |                                |
| DT  | 28-FEB-2003 (Rel. 41, Created)                                |  |            |                                |
| DT  | 28-FEB-2003 (Rel. 41, Last sequence update)                   |  |            |                                |
| DT  | 10-OCT-2003 (Rel. 42, Last annotation update)                 |  |            |                                |
| DE  | Piccolo protein (Multidomain presynaptic cytomatrix protein). |  |            |                                |

GN PCLO.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.  
RX MEDLINE=20170257; PubMed=10707984;  
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,  
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;  
RT "Piccolo, a presynaptic zinc finger protein structurally related to  
RT bassoon.";  
RL Neuron 25:203-214(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;  
RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND  
RP ALA-4694.  
RX MEDLINE=21181819; PubMed=11285225;  
RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;  
RT "An unusual C(2)-domain in the active-zone protein piccolo:  
RT implications for Ca(2+) regulation of neurotransmitter release.";  
RL EMBO J. 20:1605-1619(2001).  
CC -!- FUNCTION: May act as a scaffolding protein involved in the  
CC organization of synaptic active zones and in synaptic vesicle  
CC trafficking (By similarity).  
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic  
CC junctions.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q9JKS6-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9JKS6-2; Sequence=VSP\_003930, VSP\_003931;  
CC DOMAIN: C2 domain 1 is involved in binding calcium and  
CC phospholipids. Calcium binds with low affinity but with high  
CC specificity and induces a large conformational change.  
CC -!- SIMILARITY: Contains 2 C2 domains.  
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF138789; AAF07822.2; -.  
DR EMBL; AF227534; AAF63196.1; -.  
DR HSSP; P04410; 1A25.  
DR GO; GO:0045202; C:synaptic junction; IDA.  
DR GO; GO:0005509; F:calcium ion binding; IDA.  
DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.  
DR GO; GO:0005522; F:profilin binding; ISS.  
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.  
DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.  
DR InterPro; IPR000008; C2.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR008899; Znf\_piccolo.  
DR Pfam; PF00168; C2; 2.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF05715; Zf\_piccolo; 2.  
DR SMART; SM00239; C2; 2.  
DR SMART; SM00228; PDZ; 1.  
DR PROSITE; PS00499; C2\_DOMAIN\_1; 1.  
DR PROSITE; PS50004; C2\_DOMAIN\_2; 2.  
DR PROSITE; PS50106; PDZ; 1.

KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;  
KW Repeat; Alternative splicing.  
FT DOMAIN 372 491  
FT 12 X 10 AA TANDEM APPROXIMATE REPEATS OF  
FT P-A-K-P-Q-P-Q-P-X.  
FT C4-TYPE (POTENTIAL).  
FT C4-TYPE (POTENTIAL).  
FT POLY-PRO.  
FT PDZ.  
FT C2 DOMAIN 1.  
FT C2 DOMAIN 2.  
FT TKPTN -> SKRRK (in isoform 2).  
FT /FTid=vsp\_003930.  
FT Missing (in isoform 2).  
FT /FTid=vsp\_003931.  
FT D->A: COMPLETE LOSS OF CALCIUM-BINDING  
FT AND CALCIUM-DEPENDENT PHOSPHOLIPID  
FT BINDING ACTIVITY.  
FT D->A: COMPLETE LOSS OF CALCIUM-BINDING  
FT AND CALCIUM-DEPENDENT PHOSPHOLIPID  
FT BINDING ACTIVITY.  
FT V->S: SMALL INCREASE IN AFFINITY FOR  
FT CALCIUM.  
FT VM->SS: 10-FOLD INCREASE IN AFFINITY FOR  
FT CALCIUM.  
FT M->S: INCREASED AFFINITY FOR CALCIUM.  
FT VV->SS: 10-FOLD INCREASE IN AFFINITY FOR  
FT CALCIUM.  
FT QN->AA: MODERATE INCREASE IN AFFINITY FOR  
FT CALCIUM.  
FT A->S: NO EFFECT ON CALCIUM-BINDING  
FT ACTIVITY.  
FT SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;  
Query Match 3.2%; Score 168; DB 1; Length 5085;  
Best Local Similarity 19.6%; Pred.No.0.22;  
Matches 142; Conservative 117; Mismatches 272; Indels 192; Gaps 34;  
QY 5 PGCEPCT--KATRWGAATTSPAA---SDARSPFRQRRVLDPKDAHVQFRVPPSPACV 58  
Db 1249 PGATPLTLPKAGEKERAVAQPAEQAGSSKDGQGSKEKTEKEKSDTSSSQPKSPQGL 1308  
QY 59 PGQAGQHRGSATSLVFKQKTITSMW---TKGIK--TAESSELDSEKNNNTRIESMMSS 112  
Db 1309 SDTGYSDDGSSGLGEIPSLIPS--DEKDLKGLKDSFSQESSPSSDLAKLESTVLS 1366  
QY 113 V---QKDNFYQHNVVKLVNVSQSLDKSLTEKSTQYLNQHQTAAAMCKWQNEGKHEQLLE 169  
Db 1367 ILEAQASITLVGEKAEKKTQPKISPEKPDQKQTQTASELTITISE-----EEIKE 1418  
QY 170 SEPQTTLVPEQFSNANI-----DRSPQNDHSDTSDSENRDNQQLFTTVK 215  
Db 1419 SQEKKVS--PKDSEGGFPRKEKKEKPELVDDLSPPRASYSVDSSESSENSPVVRRKR 1476  
QY 216 LANAKQTTDEHAREAKSHQCKSKSCHPCGEDCASCQODEIDVVPKSPSLSDVGSSEVGTGS 275  
Db 1477 RTSIGSSSSDEYKQE-----DSQSGE-----EEDFIRKQ-IIEMSADEDASGS 1519  
QY 276 KNDNKLIRQE-----SCLGNPPFEKES-----EPESPMVDNKNKNSCQDS 316  
Db 1520 E-DEEFIRSQLKEISGVGESQKREEAKGKGVAGKHRLTRKSTSTSFDDAGRRHSWHD 1578  
QY 317 EADEETSPGFDEQEDGSSSQTKANKPS-----RFOARDADIEFRKRYSTKGEV 364  
Db 1579 E-DDET---FDESPELKFRETKSQSEELVVGSGGLRRFKTIELNSTIADKYSSSESSQK 1634  
QY 365 R--LHFQFE-----GGESRT---GMNDLNAK-----LPGNISSLNVCEKNSKHQ 403  
Db 1635 KTIIFYDEEPELEMEISLTDSPEDRSRGEGSSSLHASSFTPGTSPTSVSSLDSDSDSPSH 1694  
QY 404 GKXDSK-----ITDHLMLPKAEDR-----RKEQWETKHQTERKI----- 439  
Db 1695 KKGESQQRKARHRSHGPLLTPTIEDSSSEELREELKEEELLKEQEQRELEQQRKSSSKKS 1754



QY 440 -----PKYVPHLSPDKKWLGTPIEEMRRMPCGIRLPLLRPS-ANHTVT 483  
Db 1755 KKDKDELRAQRRRERPKTPPSNLSPED--ASPTIELRQAEME---ELHRSSCSSEYSPS 1809  
QY 484 IRVDLLRAGEVPKFPFTHYKDLWD-NKHVKMPCSEQNLYPVEDENGERTAGSRWELIOTA 542  
Db 1810 IESD-----PEGFEISPEKIEVKYKLPAAVSLYSPTDEQSVMQKEG-----VQKA 1857  
QY 543 LLNKFTRPQNLKDALIKYNVAYSKKWDFAL--IDFWDKVLEBAE-----AQHLYQSI 593  
Db 1858 -----LKSAEMEYEMMQPHKYKAFPAANERDEVFEKEPLYGGMLIEDYIYESL 1907  
QY 594 LPD 596  
Db 1908 VED 1910

RESULT 5  
ATRX\_PANTR STANDARD; PRT; 2492 AA.  
AC Q7YQM4;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked  
DE nuclear protein) (XNP).  
GN ATRX.  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
OX NCBI\_TaxID=9598;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22763540; PubMed=12777533;  
RA Kitano T., Schwarz C., Nickel B., Paabo S.;  
RT "Gene diversity patterns at 10 X-chromosomal loci in humans and  
RT chimpanzees";  
RL Mol. Biol. Evol. 20:1281-1289(2003).  
CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies  
CC gene expression by affecting chromatin. May be involved in brain  
CC development and facial morphogenesis.  
CC -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and  
CC phosphatidylcholine/phosphatidylserine-dependent manner (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric  
CC heterochromatin during interphase and mitosis, probably by  
CC interacting with HP1.  
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; AB102642; BAC81111.1; -.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;  
KW Zinc-finger.  
FT ZN\_FING 220 268 PHD-TYPE.  
FT NP\_BIND 1594 1601 ATP (POTENTIAL).  
FT SITE 1719 1722 DEGH BOX.  
FT DOMAIN 745 750 POLY-SER.  
FT DOMAIN 1151 1156 POLY-SER.

FT DOMAIN 1166 1169 POLY-LYS.  
FT DOMAIN 1202 1206 POLY-SER.  
FT DOMAIN 1259 1266 POLY-ASP.  
FT DOMAIN 1443 1466 POLY-GLU.  
FT DOMAIN 1499 1502 POLY-GLU.  
FT DOMAIN 1929 1939 POLY-LYS.  
FT DOMAIN 1941 1948 POLY-SER.  
FT DOMAIN 2222 2225 POLY-LYS.  
FT DOMAIN 2262 2265 POLY-GLU.  
FT DOMAIN 2420 2425 POLY-GLN.  
SQ SEQUENCE 2492 AA; 282567 MW; 56978AC7D37400DC CRC64;  
Query Match 3.1%; Score 162; DB 1; Length 2492;  
Best Local Similarity 19.1%; Pred. No. 0.19;  
Matches 203; Conservative 158; Mismatches 358; Indels 344; Gaps 55;  
QY 75 KQKITSMWMTKGIKTAESSELSKEN--NNTRIESMMSSVQKDNFYQHNVEKLVNVSQ 132  
Db 1136 ERRLSSKRNTKEIQSGSSS-DAEESSEDNKKKQQTSSKKKAVIVKEKKRNSLTSTK 1194  
QY 133 SLDKSLTEKSTOYLNQHTAAMCKWQNEGKHTQLLESEPTVTLVPEQ---FSNANIDR 189  
Db 1195 RKQADITSSSSDIEDDDQNSI---GEGSSDEQKI--KPVTENLVSSHTGFCQSSGDE 1248  
QY 190 S-----PQNDHSDTDSSENRDNQQLTIVK--LANAKQTTEDEHAREAKSHQKSKS 240  
Db 1249 ALSKSVPTVDDDDNDPENRIAKKMLLEEIKANLSSDEGSSDDEPEGKKRTGKQNE 1308  
QY 241 CHPGEDCASCQODEIDVVPKSPLDVGSSEVGTGSKND---NKLIRQESCLGNPSP-FEK 296  
Db 1309 ENFGDEEAKQVNS-----ESDSDSEE---SKPFRYRHLRLRHKLTVSDGESGEK 1356  
QY 297 ESEPESPMDV---DNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEF 353  
Db 1357 KTKPEKHKEVGRNRKVSSEDSDFQESGVSEV--SESEDEQRPRTRSKAKAELEE 1414  
QY 354 RKR-YSTKGGEVRLHFQFEGG-----ESRTGMNDLNAKLP 387  
Db 1415 NQSYKQKKRRRIKVQEDSSSENKSNSEEEEEKEEEEEEEEEEEEEDEED-DSKSP 1473  
QY 388 GNISSLNVECRNSKHGK--KDSKI-TDHLMLRPLKAEDEKKEQWETKHQRTERKIPKYVP 444  
Db 1474 G-----KGRKKIRKILKDDKLRTEQNALKEEERRKRIARERER--EKLREVIE 1522  
QY 445 -PHLSPDKKWLGTPI-----EEMRRMPCGIRLPLLRPSANHTVTIRVDLLRAGEVPKPF 498  
Db 1523 IEDASPTKCPITTKLVLDDEETKB-----PLVQVHRNMVVKL-----KP- 1562  
QY 499 PTHYKD---LWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLK 554  
Db 1563 --HQVDGVQFMWD-----CCCESV---KTKKSPSGCILAHCMLGK-----TLQ 1603  
QY 555 DALLKYNVAYSKKWDF-TALI-----DFWDKVLLEAEAOHLYQSILPDMVK 599  
Db 1604 VVSFLHTVLLCDKLDFTALVVCPLNTALNWNNEFEKQWGLKDD- 1649  
QY 600 IALCLPNICTQPIPLLKQNMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYDDI 659  
Db 1650 -KLEVSELATVKRP---QERSYMLQRWQEDGGVMIIG-----YEMY--- 1686  
QY 660 NFNRLEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFTQSLDFPEWERCEKPL---- 715  
Db 1687 --RNLAQGRNV-KSRKLKEIF--NKALVDPGPDFVVCDEGHILKN--EASAVSKAMNSIR 1739  
QY 716 TRLHVTYEGT-----IEENGQMLQVDFANRFV-----GGGVTS----- 749  
Db 1740 SRRRIILGTPLQNNLIEYHCHMVNFIKENLLGSIK-EFRNRFINPIQNGQCADSTMVDVR 1798  
QY 750 -----AGLVQEE-----IRFL-----INPELIISRLFTEVLCHNECLI 782  
Db 1799 VMKKAHILYEMLAGCVQRKDYALTATKFLPPKHEYVLAVRMTSIQCKLYQYYLDH----- 1853  
QY 783 ITGTEQYSE-----YTGyaETyRW-----SRSHEDGSRDDCERRCTE 820



```
Db 1854 LTGVGNSEGGRGKAGAKLQDFQMLSRWHPWCLQLDYISKENKGYFDEDSMD-----E 1909
QY 821 IVAIDALHFRYL--DQFVPEKMRRELKAYCGFLPGVSSSENLSAVATGN----- 869
Db 1910 FIASDSDETSLSDDYTKKKKKGKCK-----KDSSSSGSDNDVEVIKV 1957
QY 870 WCGAGAFGDARLKALIQILAAAAAERDVVYTFGDSSELMRDYSHHIFLTERKLTVDVY 929
Db 1958 WNSRSRGGGE-----GNVDETGNPNPSVSLKEESKAT----- 1989
QY 930 KLLRLRYNEECNRCSTPGPDIKLYPIYHAYESCAETADHSGQ 972
Db 1990 -----SSSNPSPAPDW----YKDFVTDAAEVLEHSGK 2019
```

```
RESULT 6
ID ATRX MOUSE STANDARD; PRT; 2476 AA.
AC Q61687;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcriptional regulator ATRX (X-linked nuclear protein)
DE (Heterochromatin protein 2) (HP1 alpha-interacting protein) (HP1-BP38
protein).
GN ATRX OR XNP OR HP1BP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98213653; PubMed=9545503;
RA Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;
RT "Comparison of the human and murine ATRX gene identifies highly
conserved, functionally important domains.";
RL Mamm. Genome 9:400-403(1998).
RN [2]
RP SEQUENCE OF 325-1176 FROM N.A.
RX MEDLINE=97133299; PubMed=8978696;
RA le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
RA Jeanmougin F., Losson R., Chambon P.;
RT "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
control of transcription by nuclear receptors.";
RL EMBO J. 15:6701-6715(1996).
```

```
[3]
RN SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE=20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATRX) at
pericentromeric heterochromatin and the short arms of acrocentric
chromosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC gene expression by affecting chromatin.
CC -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC phosphatidylcholine/phosphatidylserine-dependent manner (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC heterochromatin during interphase and mitosis, probably by
CC interacting with HP1.
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
```

```
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026032; AAC08741.1; -.
DR EMBL; X99643; CAAG7962.1; -.
DR MGD; MGI:103067; Atrx.
DR GO; GO:0000228; C:nuclear chromosome; IDA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDc; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW Zinc-finger.
FT ZN_FING 219 267 PHD-TYPE.
FT NP_BIND 1579 1586 ATP (POTENTIAL).
FT SITE 1704 1707 DEGH BOX.
FT DOMAIN 319 322 POLY-SER.
FT DOMAIN 735 738 POLY-SER.
FT DOMAIN 1001 1004 POLY-GLU.
FT DOMAIN 1130 1135 POLY-SER.
FT DOMAIN 1182 1185 POLY-SER.
FT DOMAIN 1238 1245 POLY-ASP.
FT DOMAIN 1484 1487 POLY-GLU.
FT DOMAIN 1924 1931 POLY-SER.
FT DOMAIN 2205 2208 POLY-LYS.
FT DOMAIN 2245 2248 POLY-GLU.
FT DOMAIN 2403 2408 POLY-GLN.
SQ SEQUENCE 2476 AA; 278601 MW; 90A42B790FC4FF4C CRC64;
```

```
Query Match 3.1%; Score 161; DB 1; Length 2476;
Best Local Similarity 20.2%; Pred.No.0.22;
Matches 220; Conservative 141; Mismatches 397; Indels 332; Gaps 55;

QY 53 SSPACVPGQAGQHRGSATSLVFK-----QKTTTSMWDTKGKTAESE 94
Db 1075 SLPKSPGKRPECSSSDTEKSLKGCCDSTEKRPKRIDLRRNRSSSKRNTKEVKSASS 1134

QY 95 S--LDSKENNNTNRIESMMSSVQKDNFYQHNVEKLVNVSQSLDKSLTEKSTQYLNQHTA 152
Db 1135 SDAEGSSEDNKKQKQRTSAKKTGNTKTKRNSLRATPKRKQVDITSSSSDIDGDDQNS 1194

QY 153 AMCKWQNEGK-----HT-----EQLLESEPTVTLVPEQFSNANIDRSPQ 192
Db 1195 AGEESDEQIKPVTVENLVLPSTGFCQSSGDEALS KVPATV----- 1237

QY 193 NDDHSDTDSEENRNQQLTTVLANAKQ---TTEDEHAREAKSHQKSKSCHPGEDCA 248
Db 1238 -DDDDDDNDPENRIAKMLLEEIKANLSSDESDGSSDDEPDGGGKRIKQSESPADD-G 1295

QY 249 SCQQDEIDVVKSPSLDVSGEDVGTGSKND---NKLIRQESCLGNPP-PEKESEPESPM 304
Db 1296 ELRREQLAVNQVNSDSDSSEE---SKKPYRHLRLRHLKLTLSDGESGEKPTKPKHK 1351

QY 305 DVDNSKNSCQDSEADEETSPGFDEQEDG-----SSSQTANKPSRFQARDADIEFRKR-YS 358
Db 1352 EAKGRNRRKVSSESDT---DFQESGVSEVSESEDEQRPRTSAKKALEENQRSYK 1407

QY 359 TKGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECR-----NSKHQHK----- 405
Db 1408 QKKRRRIKVQ-EDSSSENKSHSEEDKKEGDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1466

QY 406 ---KDSKI-TDHLMLPKAEDRRKEQWETKHQRTKIPKYVP-PHLSDDKKWLGTPI-- 458
Db 1467 KILKDDKLRTETQNALKEEEERRKRIARERER--EKLREVIEWIEDASPTKCPITTKLV 1524

QY 459 ---EEMRRMPRCGIRLPLLRPSANHTVTRVDLLRAGEVVPKFPPTHYKD-----LWDNKHV 511
Db 1525 DENEETKE-----PLVQVHRNMVVKL-----KP---HQVDGVQFMWD----- 1558

QY 512 KMPCEQNLYPVEDENGERTAGSRWELIQTLNKFTRPQNLKDALIKNVAYSKKQWDF- 570
```

Db 1559 ---CCCESV-----EKTCKSPGSCILAHCMGLGK-----TLQWVSFLHTVLLCDKLDIFS 1605

Qy 571 TALIDF-----WDKVLREAEQAHLQSYLPDMVKIALCLPNICTQPIPLKQKMNHSI 623

Db 1606 TALVVCPLNTALNWMNEFEK-----WQGLNDNEKLE--VSELATVKRP---QERSYML 1654

Qy 624 TMSQEQIASLLANAFCTFPRNNAKMKSEYSSYPDINENRFLPEGRSSRKPEKLTLCYF 683

Db 1655 QRWQEDGGVMIIG-----YEMY-----RNLAQGRNV-KSRKLKDIK--N 1690

Qy 684 RRVTEKKPTGLVTFTRQSLDEPFEWERCEKPL-----TRLHVTYEGT----- 725

Db 1691 KALVDPGDPFVWCDEGHILKN--EASAVSKAMNSIKSRRRIILTGTPLONNLIEYHCNVN 1748

Qy 726 -IEENGQMLQVDFANRFV-----GGGVTS-----AGLVQEE----- 756

Db 1749 FIKENLLGSIK-EFRNRFINPIQNGQCADSTMVDVRVMKKRAHILYEMLAGCVQRKDYTA 1807

Qy 757 -IRFL-----INPELIISRLFTFVLHDNECLIIITG-----TE-----QYSE 791

Db 1808 LTKFLPPKHEYVLAVRMTAIQCKLYQYVLDH-----LTGVGNSTEGGRGKAGAKLFQDFQ 1862

Qy 792 YTGVAETRW-----SRSHEDGERDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRE 844

Db 1863 MLSRIWTHPWCLOLDYISKENKGYDEDSMD-----EFIASDSDETSKLS--DEKKPK 1916

Qy 845 LNKAYCGFLRPGVSSENLSAVATGNWCGGAFGDARLKALIQILAAAAERDVVYFTFGD 904

Db 1917 GKKGKDSSSSGSGSDN-DVEVIKVMNSRSGG-----GD 1950

Qy 905 SEL--MERDIYSMHIFLTERKLTGVDVYKLLRLRYNEECNCTGPDIDIKLYPIYHAVES 962

Db 1951 GNMDDTGNNPSVSLKLDSEKTT-----STSNPSSPADW----YKDFVTDT 1992

Qy 963 CAETADHSGQ 972

Db 1993 DAEVLEHSGK 2002

RESULT 7

ATRX HUMAN

ID ATRX HUMAN STANDARD; PRT; 2492 AA.

AC P46100; P51068; Q15986; Q9H021; Q9NTS3;

DT 01-NOV-1995 (Rel. 32, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked

DE nuclear protein) (XNP) (Znf-HX).

GN ATRX OR RAD54L OR XH2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANTS SER-596;

RP GLU-740 AND SER-1860, AND VARIANTS ATR-X.

RX MEDLINE=97123494; PubMed=8968741;

RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,

RA Gibbons R.J.;

RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations

RT point to a common mechanism underlying the ATR-X syndrome.";

RL Hum. Mol. Genet. 5:1899-1907(1996).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).

RX MEDLINE=97386582; PubMed=9244431;

RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,

RA Colleaux L., Schwartz C., Fontes M.;

RT "Determination of the genomic structure of the XNP/ATRX gene encoding

RT a potential zinc finger helicase.";

RL Genomics 43:149-155(1997).

RN [3]

RP SEQUENCE FROM N.A., AND VARIANTS SER-596 AND GLU-740.

RX MEDLINE=22763540; PubMed=12777533;

RA Kitano T., Schwarz C., Nickel B., Paabo S.;

RT "Gene diversity patterns at 10 X-chromosomal loci in humans and

RT chimpanzees.";

RL Mol. Biol. Evol. 20:1281-1289(2003).

RN [4]

RP SEQUENCE OF 860-2492 FROM N.A.

RX MEDLINE=95179111; PubMed=7874112;

RA Stayton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V.,

RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,

RA Bianchi M.E., Consalez G.G.;

RT "Cloning and characterization of a new human Xql3 gene, encoding a

RT putative helicase.";

RL Hum. Mol. Genet. 3:1957-1964(1994).

RN [5]

RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.

RX MEDLINE=94214473; PubMed=8162050;

RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,

RA Millasseau P., Khrestchatsky M., Fontes M.;

RT "Cloning and expression of the murine homologue of a putative human

RT X-linked nuclear protein gene closely linked to PGK1 in Xql3.3.";

RL Hum. Mol. Genet. 3:39-44(1994).

RN [6]

RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.

RX MEDLINE=95211835; PubMed=7697714;

RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;

RT "Mutations in a putative global transcriptional regulator cause X-

RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";

RL Cell 80:837-845(1995).

RN [7]

RP SEQUENCE OF 1375-2492 FROM N.A.

RA Pearce A., Chapman J.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN [8]

RP EZH2 BINDING.

RX MEDLINE=98167853; PubMed=9499421;

RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,

RA Colleaux L.;

RT "Specific interaction between the XNP/ATR-X gene product and the SET

RT domain of the human EZH2 protein.";

RL Hum. Mol. Genet. 7:679-684(1998).

RN [9]

RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC

RP HETEROCHROMATIN.

RX MEDLINE=20040663; PubMed=10570185;

RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,

RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,

RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;

RT "Localization of a putative transcriptional regulator (ATRX) at

RT pericentromeric heterochromatin and the short arms of acrocentric

RT chromosomes.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).

RN [10]

RP DISEASE.

RX MEDLINE=20213147; PubMed=10751095;

RA Villard L., Fontes M., Ades L.C., Gecz J.;

RT "Identification of a mutation in the XNP/ATR-X gene in a family

RT reported as Smith-Fineman-Myers syndrome.";

RL Am. J. Med. Genet. 91:83-85(2000).

RN [11]

RP DISEASE.

RX MEDLINE=22804448; PubMed=12858175;

RA Gibbons R.J., Pellagatti A., Garrick D., Wood W.G., Malik N.,

RA Ayyub H., Langford C., Boulwood J., Wainscoat J.S., Higgs D.R.;

RT "Identification of acquired somatic mutations in the gene encoding

RT chromatin-remodeling factor ATRX in the alpha-thalassemia

RT myelodysplasia syndrome (ATMDS).";

RL Nat. Genet. 34:446-449(2003).

RN [12]

RP VARIANT ATR-X SER-1713.

RX MEDLINE=97196774; PubMed=9043863;

RA Villard L., Lacombe D., Fontes M.;

RT "A point mutation in the XNP gene, associated with an ATR-X phenotype

RT without alpha-thalassemia.";

RL Eur. J. Hum. Genet. 4:316-320(1996).  
RN [13]  
RP VARIANT JM GLN-2131.  
RX MEDLINE=96224392; PubMed=8630485;  
RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,  
RA Munnich A., Lyonnet S.;  
RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";  
RL Nat. Genet. 12:359-360(1996).  
RN [14]  
RP VARIANTS ATR-X.  
RX MEDLINE=97467722; PubMed=9326931;  
RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,  
RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,  
RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,  
RA Higgs D.R.;  
RT "Mutations in transcriptional regulator ATRX establish the functional  
RT significance of a PHD-like domain."  
RL Nat. Genet. 17:146-148(1997).  
RN [15]  
RP VARIANT ATR-X LEU-246.  
RX MEDLINE=20123062; PubMed=10660327;  
RA Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,  
RA Greco D., Cardoso C., Fontes M., Ragusa A.;  
RT "New mutations in XNP/ATR-X gene: a further contribution to  
RT genotype/phenotype relationship in ATR/X syndrome."  
RL Hum. Mutat. 12:214-214(1998).  
RN [16]  
RP VARIANT SHS LYS-1742.  
RX MEDLINE=99347960; PubMed=10417298;  
RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,  
RA Prieto F., Fontes M., Martinez F.;  
RT "Mutation of the XNP/ATR-X gene in a family with severe mental  
RT retardation, spastic paraplegia and skewed pattern of X inactivation:  
RT demonstration that the mutation is involved in the inactivation  
RT bias."  
RL Am. J. Hum. Genet. 65:558-562(1999).  
RN [17]  
RP VARIANT CWS THR-2050.  
RX MEDLINE=99326061; PubMed=10398237;  
RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,  
RA Curtis M.;  
RT "Carpenter-Waziri syndrome results from a mutation in XNP."  
RL Am. J. Med. Genet. 85:249-251(1999).  
RN [18]  
RP VARIANTS ATR-X GLU-175; 178-VAL--LYS-198 DEL; SER-190; PRO-219;  
RX LEU-246 AND CYS-249.  
RN MEDLINE=99219535; PubMed=10204841;  
RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Beloungue J.,  
RA Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,  
RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;  
RT "Evaluation of a mutation screening strategy for sporadic cases of  
RT ATR-X syndrome."  
RL J. Med. Genet. 36:183-186(1999).  
RN [19]  
RP VARIANTS ATR-X SER-179; LEU-190; ILE-194; CYS-246; PHE-1552; SER-1645  
RX AND CYS-1847.  
RN MEDLINE=20451413; PubMed=10995512;  
RA Wada T., Kubota T., Fukushima Y., Saitoh S.;  
RT "Molecular genetic study of Japanese patients with X-linked alpha-  
RT thalassemia/mental retardation syndrome (ATR-X).";  
RL Am. J. Med. Genet. 94:242-248(2000).  
RN [20]  
RP VARIANT ATR-X MET-1621.  
RX MEDLINE=22108790; PubMed=12116232;  
RA yntema H.G., Poppelaars F.A., Derksen E., Oudakker A.R.,  
RA van Roosmalen T., Jacobs A., Obbema H., Brunner H.G., Hamel B.C.J.,  
RA van Bokhoven H.;  
RT "Expanding phenotype of XNP mutations: mild to moderate mental  
RT retardation."  
RL Am. J. Med. Genet. 110:243-247(2002).  
CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies  
CC gene expression by affecting chromatin. May be involved in brain  
CC development and facial morphogenesis.

CC -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and  
CC phosphatidylcholine/phosphatidylserine-dependent manner (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric  
CC heterochromatin during interphase and mitosis, probably by  
CC interacting with HPI.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=5;  
CC Name=4;  
CC IsoId=P46100-1; Sequence=Displayed;  
CC Name=1;  
CC IsoId=P46100-2; Sequence=VSP\_000575;  
CC Name=2;  
CC IsoId=P46100-3; Sequence=VSP\_000574;  
CC Name=3;  
CC IsoId=P46100-4; Sequence=VSP\_000576;  
CC Name=5;  
CC IsoId=P46100-5; Sequence=VSP\_000574, VSP\_000576;  
CC TISSUE SPECIFICITY: Ubiquitous.  
CC -!- DISEASE: Defects in ATRX are the cause of X-linked alpha-  
CC thalassemia/mental retardation syndrome (ATR-X) [MIM:301040]. ATR-  
CC X is an X-linked disorder comprising severe psychomotor  
CC  
Query Match 3.1%; Score 161; DB 1; Length 2492;  
Best Local Similarity 19.1%; Pred. No. 0.22;  
Matches 203; Conservative 158; Mismatches 358; Indels 344; Gaps 55;  
QY 75 KQKITSMMDTKGIKTAESSELDSEN--NNFRIESMSSVQKNFYQHNVKLVNVSOL 132  
Db ERRNLSSKRNTKBIQSGSSSS--DAEESSEDNKKKQRTSSKKGAVIVKEKKRNSLRTSTK 1194  
QY 133 SLDKSLTEKSTOYLNQHQTAAMCKWQNEGHTEQLLESEPQTVTLVPEQ---FSNANIDR 189  
Db RKQADITSSSSSDIEDDDQNSI---GEGSSDEQKI--KPVTENLVLSSTGTCQSSGDE 1248  
QY 190 S-----PQNDHSDTDSENRDNQQFLTIVK--LANAKQTTEDEHAREAKSHQCSKS 240  
Db ALSKSVPTVDDDDNDPENRIAKKMLLEIKANLSSDEGSSDDEPEEGKRTGKQNE 1308  
QY 241 CHPGEDCASCQQDEIDVVPKSPLSVGSSEVGTGSKND---NKLIRQESCLGNPP--FEK 296  
Db ENPGDEEAKQVNS-----ESDSDEE---SKKPRYRHLRLRHKLTVSDGESGEK 1356  
QY 297 ESEPESPMDV---DNSKNSCQDSEADEETSPGDEQEDGSSSQTANKPSRFQARDADIEF 353  
Db KTKPKEHKEVKGRRNRKVSSDESDSDFOESGVSEEV--SESEDEQRPRTRSKAKALEEE 1414  
QY 354 RKR-YSTKGGEVRLHFPQEGG-----ESRTGMNDLNAKL 387  
Db NQRSYKQKKRRRIKVKQEDSSSENKSNSEEEEEKEEEEEEEEEEEEEDEND--DSKSP 1473  
QY 388 GNISSLNVECRNSKQHGK--KDSKI--TDHLMRLPKAEDRRRKEQWETKHQRTERKIPKVP 444  
Db G-----KGRKIRKILKDDKLRRTETQNALKEEERKRIRAEERER--EKLREVIE 1522  
QY 445 -PHLSPDKKWLGTPI-----EEMRRMPCGIRLLPLLRPSANHTVTIRVDLLRAGEVPKPF 498  
Db IEDASPTKCPITTKLVLDDEETKE-----PLVQVHRNMVIKL-----KP- 1562  
QY 499 PTHYKD---LWDNKHVMPKSEONLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLK 554  
Db --HQVDGVQFMWD-----CCCESV-----KTKKSPGSGCILAHCMLGK-----TLQ 1603  
QY 555 DAILKYNVAYSKKWF-TALI-----DFWDKVLSEAEQAHLYSILPDMVK 599  
Db VVSFLHTVLLCDKLDFTALVVCPLNTALNMWMEFEKQEGKDD----- 1649  
QY 600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLIANAFCTFPRRNAKMKSEYSSVPI 659  
Db -KLEVSELATVVRP---QERSYMLQRWQEDGGVMIIG-----YEMY--- 1686  
QY 660 NFNLFEGRSSRKPEKLTLCFYFRRVTEKKPTGLVTFRQSLDFPEWERCEKPL----- 715



Db 1687 --RNLAQGRNV-KSRKLKEIF--NKALVDPGPDFVVCDEGHILKN--EASAVSKAMNSIR 1739

QY 716 TRLHVTEGT-----IEENGQMLQVDFANRFV-----GGVTS----- 749

Db 1740 SRRRIILGTPLQNNLIEYHCVNFIKENLLGSIK-EFRNRFINPIQNGQCADSTMVDVR 1798

QY 750 -----AGLVQEE-----IRFL-----INPELIISRLTFTEVLDHNECLI 782

Db 1799 VMKKRAHILYEMLAGCVQRKDYALTATKFLPPKHEYVLAVRMTSIQCKLYQYYLDH----- 1853

QY 783 ITGTEQYSE-----YTGVAETYRW-----SRSHEDGSEDDCERRCTE 820

Db 1854 LTGVGNSEGGRGKAGAKLFQDFQMLSRITWHPWCLQLDYISKENKGYFDEDSMD-----E 1909

QY 821 IVALDALHFRYL--DQFVPEKMRRELNKAYCGFLRPGVSSENLSAVATGN----- 869

Db 1910 FIASDDETSMSLSDDYTKKKKKKKGK-----KDSSSSGSGSDNDVEVIKV 1957

QY 870 WGCAGFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTGVGDVY 929

Db 1958 WNSRSGGGE-----GNVDETGNPVSLSKLEESKAT----- 1989

QY 930 KLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSQ 972

Db 1990 -----SSNPSPAPDW----YKDFVTDADAEVLEHSGK 2019

RESULT 8

ATRX\_PONPY STANDARD; PRT; 2492 AA.

ID ATRX\_PONPY Q7YQM3;

AC Q7YQM3;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked nuclear protein) (XNP).

GN ATRX.

OS Pongo pygmaeus (Orangutan).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.

OX NCBI\_TaxID=9600;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22763540; PubMed=12777533;

RA Kitano T., Schwarz C., Nickel B., Paabo S.;

RT "Gene diversity patterns at 10 X-chromosomal loci in humans and chimpanzees";

RL Mol. Biol. Evol. 20:1281-1289(2003).

CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies gene expression by affecting chromatin. May be involved in brain development and facial morphogenesis.

CC -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and phosphatidylcholine/phosphatidylserine-dependent manner (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric heterochromatin during interphase and mitosis, probably by interacting with HPI.

CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.

CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AB102643; BAC81112.1; --

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR000330; SNF2\_N.

DR Pfam; PF00271; helicase\_C\_1.

DR Pfam; PF00176; SNF2\_N; 1.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELICC; 1.

KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding; Zinc-finger.

KW ZN\_FING 220 268 PHD-TYPE.

FT NP\_BIND 1594 1601 ATP (POTENTIAL).

FT SITE 1719 1722 DEGH BOX.

FT DOMAIN 745 750 POLY-SER.

FT DOMAIN 1151 1156 POLY-SER.

FT DOMAIN 1166 1169 POLY-LYS.

FT DOMAIN 1202 1206 POLY-SER.

FT DOMAIN 1259 1266 POLY-ASP.

FT DOMAIN 1443 1466 POLY-GLU.

FT DOMAIN 1499 1502 POLY-GLU.

FT DOMAIN 1929 1939 POLY-LYS.

FT DOMAIN 1941 1948 POLY-SER.

FT DOMAIN 2222 2225 POLY-LYS.

FT DOMAIN 2262 2265 POLY-GLU.

FT DOMAIN 2420 2425 POLY-GLN.

SQ SEQUENCE 2492 AA; 282614 MW; 48EC97106D0CE20 CRC64;

Query Match 3.1%; Score 161; DB 1; Length 2492;

Best Local Similarity 19.1%; Pred. No. 0.22;

Matches 203; Conservative 158; Mismatches 358; Indels 344; Gaps 55;

QY 75 KQKTTTSMWDTGKGTAESESLDSKEN--NNTRIESMMSVQKDNFYQHNVKLVNVSQ 132

Db 1136 ERRNLSSKRNTKEIQSGSSS--DAEESSEDNKKKQRTSSKKAVIVKEKKRNSLRTSTK 1194

QY 133 SLDKSLTEKSTQYLNQHTAAMCKWQNEGKHTQLLESEPOTVTLVPEQ---FSNANIDR 189

Db 1195 RKQADITSSSSSDIEDDDQNSI---GEGSSDEQKI--KPVTEINLVLSSTGFCQSSGDE 1248

QY 190 S-----PQNDHSDTDSENRDNQQFLTTVK--LANAKOTTEDEHAREAKSHQKCSKS 240

Db 1249 ALSKSVPTVVDNDDDDNDPENRIAKKMLLEEIKANLSSDEGSSDDEPEEGKRTGKQNE 1308

QY 241 CHPGEDCASCQQDEIDVVPKSPISDVSGSESDVGTGSKND---NKLIRQESCLGNPSP-FEK 296

Db 1309 ENPGDEEAKQVNS-----ESDSDSEE---SKKPRYRHRLLRHKLTVSDGESGEEK 1356

QY 297 ESEPESPMDV---DNSKNSCODSEADEETS PGDFDEQEDGSSSOTANKPSRFQARDADIEF 353

Db 1357 KTKPEHKEVGNRRRKVSSSESDSDFQESGVSEEV--SESEDEQRPRTSRAKKALEE 1414

QY 354 RKR-YSTKGGEVRLHFQEGG-----ESRTGMNDLNAKLP 387

Db 1415 NQRTYKQKKRRRIKQVEDSSSENKSNSEEEEEEEEEEEEEEEEEEDEND-DSKSP 1473

QY 388 GNISSLNVECRNSKHGK--KDSKI-TDHLMLPLKAEADRRKQWETKHQTERKIPKYVP 444

Db 1474 G-----KGRKKIRKILKDDKLTETQNALKEEEERRKRAERER--EKLREVIE 1522

QY 445 -PHLSPDKKWLGTPI-----ENMRMRPCGIRLPLLRPSANHTVTIRVDLLRAGEVPKPF 498

Db 1523 IEDASPTKCPITTKLVLDEDEETKE-----PLVQVHRNMVVKL-----XP- 1562

QY 499 PTHYKD---LWDNKHVKMPCSEONLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLK 554

Db 1563 --HQVDGVQFMWD-----CCCESV-----KTKKSPGSGCILAHCMLGK-----TLQ 1603

QY 555 DAILKVNVAYSKKWDF-TALI-----DFWDKVLBEAAEAQHLYSILPDMVK 599

Db 1604 VVSFLHTVLLCDKLDFTALVVCPLNTALNMWNEFEKQWQGLKDE----- 1649

QY 600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRNAKMKSEYSSYPDI 659

Db 1650 -KLEVSELATVKRP---QERSYMLQRWQEDGGVMIIG-----YEMY--- 1686

QY 660 NFNRLFEGRSRRKPEKLTLCFYFRRVTEKKPTGLVTFTRQSLDEDFPEWERCEKPL---- 715

Db 1687 --RNLAQGRNV-KSRKLKEIF--NKALVDPGPDFVVCDEGHILKN--EASAVSKAMNSIR 1739



QY 716 TRLVHTYEGT-----IEENGQMLQVDFANRFV-----GGGVTs----- 749  
Db 1740 SRRRIILTGTPLQNNLIEYHCVNFIKENLLGSIK-EFRNRFINPIQNGQCADSTMVDVR 1798  
QY 750 -----AGLVQEE-----IRFL-----INPELIISRLFTFVLHDHNECLI 782  
Db 1799 VMKKAHILYEMLAGCVQRKDYALTATKFLPPKHVEYLAVRMTSIQCKLYQYYLDH----- 1853  
QY 783 ITGTEQYSE-----YTYGAETVRW-----SRSHEDGSRDDCERRCTE 820  
Db 1854 LTGVGNNSGGRGKAGAKLFQDFQMLSRITWHPWCLQLDYISKENKGYFDEDSMD-----E 1909  
QY 821 IVALDALHFRYL--DOFVPEKMRRELNKAYCGFLRPGVSSENLSAVATGN----- 869  
Db 1910 FIASDSDETSMSLSSDDYTKKKKKKKGK-----KSSSSSGSGSDNDVEVIKV 1957  
QY 870 WCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTVDVY 929  
Db 1958 WNSRSRGGGE-----GNVDETGNPNPSVSLKLEESKAT----- 1989  
QY 930 KLLRLRYNEECRNCSTGPDIKLYPIYHAYESCAETADHSGQ 972  
Db 1990 -----SSSNPSSPADW-----YKDFVTDADAELVLEHSGK 2019

RESULT 9  
BMS1\_YEAST STANDARD; PRT; 1183 AA.  
AC Q08965;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ribosome biogenesis protein BMS1.  
GN BMS1 OR YPL217C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RX MEDLINE=97313271; PubMed=9169875;  
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Anson W.,  
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,  
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hebling U., Heumann K., Hilbert H., Kaiman S., Kleine K.,  
RA Hunnicke-Smith S., Hyman R., Johnston M., Kaiman S., Kleine K.,  
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,  
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,  
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,  
RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,  
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,  
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,  
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";  
RL Nature 387:103-105(1997).  
RN [2]  
RP FUNCTION.  
RX MEDLINE=21449426; PubMed=11565748;  
RA Wegierski T., Billy E., Nasr F., Filipowicz W.;  
RT "Bmslp, a G-domain-containing protein, associates with Rcl1p and is  
RT required for 18S rRNA biogenesis in yeast."  
RL RNA 7:1254-1267(2001).  
CC -!- FUNCTION: May act as a molecular switch during maturation of the  
CC 40S ribosomal subunit in the nucleolus. The depletion of BMS1  
CC interferes with processing of the 35S pre-rRNA at sites A0, A1,  
CC and A2, and the formation of 40S subunits.  
CC -!- SUBUNIT: Associates with RCL1.

CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -!- SIMILARITY: Belongs to the BMS1 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z73573; CAA97932.1; --  
DR PIR; S65236; S65236.  
DR SGD; S0006138; BMS1.  
DR GO; GO:0005737; C:cytoplasm; IDA.  
DR GO; GO:0005730; C:nucleolus; IDA.  
DR GO; GO:0006365; P:35S primary transcript processing; IDA.  
DR GO; GO:0042255; P:ribosome assembly; IDA.  
DR InterPro; IPR003593; AAA ATPase.  
DR InterPro; IPR007034; DUF663.  
DR Pfam; PF04950; DUF663; 1.  
DR SMART; SM00382; AAA; 1.  
KW Ribosome biogenesis; Nuclear protein; ATP-binding.  
FT NP\_BIND 76 83 ATP (POTENTIAL).  
SQ SEQUENCE 1183 AA; 135570 MW; 9A337FIEE0B0F21D CRC64;  
  
Query Match 3.1%; Score 159.5; DB 1; Length 1183;  
Best Local Similarity 18.3%; Pred. No. 0.1;  
Matches 156; Conservative 116; Mismatches 274; Indels 307; Gaps 36;  
  
QY 33 SRQRRVLDPKD-----AHVQFRVPSPSPACVPGQ-AGQHRGSATSL 72  
Db 341 TRRRRLDDKDLIYAPMSDVGGVLMKDVAVIDIGKNEEPSFVPGQERGEKLMTGL 400  
QY 73 VFQKQTTITSMMDTKGI-----KTAESSELSKNNNTRIE-----SMSSSVQKD 116  
Db 401 QSVEQSIAEKFDGVGLQLFNSGTTELHEVADHEGMD-VESGEESIEDDEGSKGRTSLRKP 459  
QY 117 NFYQHNV-EKLVNVSQSLDKSLTEKSTQYLNHQHTAAMCKWQNEGKHTEQLLESEPTV 175  
Db 460 RIYKPVQEEADADIDNLPSE-----PYTND-----DDVQDSEPRMV 497  
QY 176 TLVPEQFSNANIDRSPQNDHSDTSDSENRDNQOFLTIVKLANAKQTTEDEHA----- 228  
Db 498 EI--DFNNTGEQGAELALETDFEESSEDEFWERT--ANKLKTESKRTWNIGKL 552  
QY 229 --REAKSHQKCSCHPGEDCASCQODEIDVVPKSLSDVGSADV-----GTGSKN 277  
Db 553 IYMDNISPEECIRWR-GEHDDSD--KDESDI-----EEDVDDDFRKKDGTVKE 599  
QY 278 DNK--LIRQESCLGNPPFEKESPESEPMDV-----DNSKNSCQDSEADEETS 323  
Db 600 GNKDHAVDLEKVFYDFTFEKLAKKWSVDALKERFLGAGILGNDNKTSDSNEGGEELY 659  
QY 324 PGFDEQEDGSSSQTANKPSRFQARDAD-----IEFRKY 357  
Db 660 GDFEDLEDGNPSEQAEDNSDKSEDEDEDEDNEDTNGDDDDNSFTNDAEKKDLTNEQERMN 719  
QY 358 STKGGEVRLHFQFEGGESRTGMNDLN-----AKLPGNISSLNVEC-----RNS 400  
Db 720 AAKKEKLAQFEIBEGENFKEDDENNEYDTWYELQAKISKOLEINNIEYQEMTFEQRQR 779  
QY 401 KQHGKKDSKITDHLMLRP-----KAEDRRKEQWE--- 429  
Db 780 IEGFKAGSYVRIVFEKVPMEFVKNFNPKFPIVMGGLLPTKFGIVKARLR--HRWHKKI 838  
QY 430 -----TKHQRTKIPKVPVPPHLSPPDKKWLK---TPIEMRR 463  
Db 839 LKTNPLVLSLGRFRFQTLPIYTTTDSRTFRMLKYTPHTYCYNAAFYGLPCSP-----N 893  
QY 464 MPRCGIRLPLRPSAN-----HTVTIRVDLLRAGEVPKFPFTHYKDLWDNKHVK- 512  
Db 894 TPFCGVQIVANSOTGNGFRIAATGIVEEIDVNIIEIVKKLKL-VGFP--YKIFKNTAFIKD 950

















Search completed: May 26, 2004, 18:41:09  
Job time : 20.6372 secs

|    |          |         |           |                                |
|----|----------|---------|-----------|--------------------------------|
| FT | CHAIN    | 21      | 646       | SECRETOTRANIN I.               |
| FT | PEPTIDE  | 418     | 484       | GAWK PEPTIDE.                  |
| FT | PEPTIDE  | 634     | 646       | SECRETOLYTIN.                  |
| FT | DISULFID | 36      | 57        | BY SIMILARITY.                 |
| FT | MOD_RES  | 158     | 158       | SULFATION (POTENTIAL).         |
| FT | MOD_RES  | 315     | 315       | SULFATION (BY SIMILARITY).     |
| FT | CONFLICT | 64      | 64        | N -> S (IN REF. 1).            |
| FT | CONFLICT | 70      | 70        | N -> D (IN REF. 2).            |
| FT | CONFLICT | 93      | 98        | SEAPGL -> FRSPRAS (IN REF. 3). |
| FT | CONFLICT | 181     | 181       | T -> M (IN REF. 2).            |
| FT | CONFLICT | 261     | 261       | H -> R (IN REF. 2).            |
| FT | CONFLICT | 386     | 386       | P -> R (IN REF. 2).            |
| FT | CONFLICT | 481     | 481       | H -> L (IN REF. 3).            |
| FT | CONFLICT | 597     | 597       | M -> V (IN REF. 2).            |
| SQ | SEQUENCE | 646 AA; | 73339 MW; | 420DB1178FD9E415 CRC64;        |

Query Match

Best Local Similarity

Matches 144; Conservative 108; Mismatches 268; Indels 225; Gaps 36;

3.0%; Score 154.5; DB 1; Length 646;

19.3%; Pred. No. 0.089;

|    |    |  |          |     |
|----|----|--|----------|-----|
| QY | 43 | DAHVOFRVPPSSPACVPGQAGQHRGSATSLVFKQTKTITSWMDTKGIK----       | TAESSELD | 98  |
| Db | 43 | NALLKSSAPITPEC--RQVLKKGKELKNEEKSENENTRFEVRLLRDPADTSEAPGLSS |          | 100 |

|    |     |  |           |     |
|----|-----|--|-----------|-----|
| QY | 99  | KENNNTRIESMWSVQKDNFYQHNVKLVNVSQLSLDKSLTEKSTQVYLNQHQTAA | MCKWQ     | 158 |
| Db | 101 | REDSGEG-DAQVPTVADTESGGHSRERAGBPPGSQVAKETRYKSEGO        | NREEMVKYQ | 159 |

|    |     |   |        |     |
|----|-----|---|--------|-----|
| QY | 159 | NEGK---HTEQLLESEP---QTVTLVPEQFSNANIDRSPQNDHSDTDSEENRD | NQOFLT | 212 |
| Db | 160 | KRERGEVGESEERLSEGPQKAQTAFL-----NORNQTPAKKBEVLSRYDTQ   | SAR    | 207 |

|    |     |  |        |     |
|----|-----|--|--------|-----|
| QY | 213 | TVKLANAKQTTDEHAREAKSHQCKSKC--HP-GEDCASQOQDEIDVVPKSP    | LSDVGE | 269 |
| Db | 208 | GLEKSHSRERSQSEETKSQENWPQELQRHPEGOEAPG--ESEEDASPEV----- |        | 257 |

|    |     |  |         |     |
|----|-----|--|---------|-----|
| QY | 270 | DVGTGKNDNKLIRQESCLGNSPPFEKESEPESPMDVDNSKNSCQDSHADEET | SPGFDEQ | 329 |
| Db | 258 | -----DKRHSRPRHHGRSRPDRSSQEGNPPL-----EESHVG----       |         | 291 |

|    |     |   |        |     |
|----|-----|---|--------|-----|
| QY | 330 | EDGSSSQTANKPSRFQARDADIEFRKRYSTKGGEVRLHFQEGGESRIGMND | LNAKLP | 389 |
| Db | 292 | TGNSDEEKARHPAHPRALEEGAEY-----GEEVRRHSA-----AQAP     | GD     | 330 |

|    |     |  |          |     |
|----|-----|--|----------|-----|
| QY | 390 | ISSLNVECRNSKHGKKDSKITDHLMLRPKAEDRRKEQWETKHQRTERKIP | KYVPPHLS | 449 |
| Db | 331 | LQGARFGGRGRGEH-----QALRRPSEES--LEQENKRH-----GLSP   |          | 366 |

|    |     |   |        |     |
|----|-----|---|--------|-----|
| QY | 450 | DKKWLGTPIEM--RRMPCRGIRLPLLR-----PSANHTVTIRVDLLR-AGE | VPKPFP | 502 |
| Db | 367 | DLNMAQGYSESEEEERGPAPG---PSYRARGGEAAAYSTLGQTDKRF     | LG     | 417 |

|    |     |   |          |     |
|----|-----|---|----------|-----|
| QY | 503 | K---DLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALNKF | TRPQ-NLK | 558 |
| Db | 418 | RVQESQDKARRRLPGELRNYLDYGEKEGEAAARGKWQ-----PQGD  | PRDA--   | 463 |

|    |     |  |         |     |
|----|-----|--|---------|-----|
| QY | 559 | KYNVAYSKKWDFTALIDPDKVLEEAQAHLVQSILPDMVKIALCLPNICTQ | PIPLLKQ | 618 |
| Db | 464 | -----DENREARLRG-----KQY                            |         | 477 |

|    |     |   |         |     |
|----|-----|---|---------|-----|
| QY | 619 | MNHSITMSQEQIASLLANAFF-----CTFPRR-----NAKMKSEYSSYP | D       | 658 |
| Db | 478 | APHHIT--EKRLGELL-NPFYDPSQWKSSRFRKDPMDDSFLEGEENGLT | NEKNFPE | 534 |

|    |     |   |         |     |
|----|-----|---|---------|-----|
| QY | 659 | INFN-----RLFE-----GRSSRKPEKLTLCYFRVRYTEKKPTGLVT   | FRQSL   | 709 |
| Db | 535 | YNYDWWWEKPFEEVNVWGYEKRNVVPKLDLKRQYDRVAELDQ--LLHYR | KKSAB-F | 591 |

|    |     |                          |   |     |
|----|-----|--------------------------|---|-----|
| QY | 710 | RCEKPLTRLHVTYEGTIEENGQGM | L | 734 |
| Db | 592 | DSEEQMSPQH-TAENEEKAGQGV  | L | 615 |



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 34.3712 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: " US-09-302-812-4\_COPY\_677\_977  
Perfect score: 1597  
Sequence: 1 KTLFCYFRVTEKKPTGLVT.....YHAVESCAETADHSGQRTGT 300

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 1597  | 100.0       | 976    | 5     | Aae25630 Human pol |
| 2          | 1597  | 100.0       | 976    | 5     | Aau76021 Human pol |
| 3          | 1597  | 100.0       | 976    | 5     | Aau76012 Human pol |
| 4          | 1597  | 100.0       | 976    | 6     | Abg72279 Human pol |
| 5          | 1497  | 93.7        | 968    | 5     | Aae25631 Murine po |
| 6          | 1497  | 93.7        | 968    | 5     | Aau76022 Mouse pol |
| 7          | 1497  | 93.7        | 968    | 5     | Aau76013 Mouse pol |
| 8          | 1497  | 93.7        | 968    | 6     | Abg72280 Murine p  |
| 9          | 1493  | 93.5        | 977    | 5     | Aae25629 Bovine po |
| 10         | 1493  | 93.5        | 977    | 5     | Aau76020 Bovine po |
| 11         | 1493  | 93.5        | 977    | 5     | Aau75799 Bovine po |
| 12         | 1493  | 93.5        | 977    | 6     | Abg72278 Bovine po |
| 13         | 619.5 | 38.8        | 768    | 4     | Abb59491 Drosophil |
| 14         | 619.5 | 38.8        | 768    | 5     | Aae25632 Fruit fly |
| 15         | 619.5 | 38.8        | 768    | 5     | Aau76023 Fruit fly |
| 16         | 619.5 | 38.8        | 768    | 5     | Aau76014 Fruit fly |
| 17         | 619.5 | 38.8        | 768    | 6     | Abg72281 Fruit fly |
| 18         | 371.5 | 23.3        | 726    | 5     | Aae25633 Poly aden |
| 19         | 371.5 | 23.3        | 726    | 5     | Aau76024 Worm poly |
| 20         | 371.5 | 23.3        | 726    | 5     | Aau76015 Worm poly |
| 21         | 371.5 | 23.3        | 726    | 6     | Abg72282 C. elegan |
| 22         | 344.5 | 21.6        | 819    | 4     | Abg20721 Novel hum |
| 23         | 210   | 13.1        | 100    | 4     | Abg11103 Novel hum |
| 24         | 178   | 11.1        | 33     | 5     | Aae25652 Bovine po |
| 25         | 167   | 10.5        | 31     | 5     | Aae25634 Bovine po |

|    |      |      |     |   |          |           |
|----|------|------|-----|---|----------|-----------|
| 26 | 167  | 10.5 | 31  | 5 | Aae25651 | Bovine po |
| 27 | 167  | 10.5 | 31  | 5 | Aau76025 | Bovine po |
| 28 | 167  | 10.5 | 31  | 5 | Aau76016 | Bovine po |
| 29 | 167  | 10.5 | 31  | 6 | Abg72283 | Oligopept |
| 30 | 149  | 9.3  | 29  | 5 | Aae25635 | Bovine po |
| 31 | 149  | 9.3  | 29  | 5 | Aau76026 | Bovine po |
| 32 | 149  | 9.3  | 29  | 5 | Aau76017 | Bovine po |
| 33 | 149  | 9.3  | 29  | 6 | Abg72284 | Oligopept |
| 34 | 97.5 | 6.1  | 395 | 6 | Abu23332 | Protein e |
| 35 | 92.5 | 5.8  | 754 | 4 | Abg10064 | Novel hum |
| 36 | 92.5 | 5.8  | 754 | 4 | Abg18445 | Novel hum |
| 37 | 92.5 | 5.8  | 754 | 4 | Abg19877 | Novel hum |
| 38 | 92.5 | 5.8  | 754 | 4 | Abg14429 | Novel hum |
| 39 | 92.5 | 5.8  | 754 | 4 | Abg02294 | Novel hum |
| 40 | 92.5 | 5.8  | 754 | 4 | Abg10211 | Novel hum |
| 41 | 92.5 | 5.8  | 754 | 4 | Abg13269 | Novel hum |
| 42 | 92.5 | 5.8  | 754 | 4 | Abg10377 | Novel hum |
| 43 | 92.5 | 5.8  | 754 | 4 | Abg12499 | Novel hum |
| 44 | 92.5 | 5.8  | 754 | 4 | Abg07848 | Novel hum |
| 45 | 92.5 | 5.8  | 754 | 4 | Abg05427 | Novel hum |

ALIGNMENTS

RESULT 1  
AAE25630  
ID AAE25630 standard; protein; 976 AA.

XX AC AAE25630;

XX DT 04-NOV-2002 (first entry)

XX DE Human poly adenosine diphosphate-ribose glycohydrolase (PARG).

XX KW Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.

XX OS Homo sapiens.

XX PN US6395543-B1.

XX PD 28-MAY-2002.

XX PF 23-FEB-2000; 2000US-00511507.

XX PR 01-MAY-1998; 98US-0083768P.

XX PR 30-APR-1999; 99US-00302812.

XX PA (KENT ) UNIV KENTUCKY RES FOUND.

XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;

XX DR WPI; 2002-535641/57.

XX DR N-PSDB; AAD42082.

XX PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX PS Claim 3; Col 55-60; 77pp; English.

XX CC The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the

CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is human PARG  
XX  
SQ Sequence 976 AA;

Query Match 100.0%; Score 1597; DB 5; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.1e-166;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFPEWERCCKPLTRLHLVYEGTIEENGQMLQV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTTQSLDFPEWERCCKPLTRLHLVYEGTIEENGQMLQV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIIITGTEQYSEYTGVA 120  
Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIIITGTEQYSEYTGVA 796  
QY 121 ETYRWSRSHEDGSDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 857 VSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 916  
QY 241 FLTERKLTVDVYKLLRLRYNEECRCNCTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTVDVYKLLRLRYNEECRCNCTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976

RESULT 2  
AAU76021  
ID AAU76021 standard; protein; 976 AA.  
XX  
AC AAU76021;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurotoxicity; Huntington's disease;  
KW Huntington's disease; Parkinson's disease.

XX Homo sapiens.  
OS  
XX US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
PA

XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-163240/21.  
DR N-PSDB; ABK14932.  
XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 2; Col 55-60; 81pp; English.  
PS  
XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the human PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention

XX SQ Sequence 976 AA;  
Query Match 100.0%; Score 1597; DB 5; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.1e-166;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFPEWERCCKPLTRLHLVYEGTIEENGQMLQV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTTQSLDFPEWERCCKPLTRLHLVYEGTIEENGQMLQV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIIITGTEQYSEYTGVA 120  
Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIIITGTEQYSEYTGVA 796  
QY 121 ETYRWSRSHEDGSDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 857 VSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 916  
QY 241 FLTERKLTVDVYKLLRLRYNEECRCNCTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTVDVYKLLRLRYNEECRCNCTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976

RESULT 3  
AAU76012  
ID AAU76012 standard; protein; 976 AA.  
XX  
AC AAU76012;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX  
OS Homo sapiens.  
XX  
XX US6333148-B1.  
XX  
PD 25-DEC-2001.  
XX  
XX 30-APR-1999; 99US-00302812.  
PF  
XX 01-MAY-1998; 98US-0083768P.  
PR  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
PA

XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-153820/20.  
XX N-PSDB; ABK14494.  
DR Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX Claim 3; Col 55-60; 80pp; English.  
PS The present invention relates to a new method for screening compounds for  
XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the human PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX  
SQ Sequence 976 AA;  
Query Match 100.0%; Score 1597; DB 5; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.1e-166;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFFPWERCEKPLTRLHVTVEGTIEENGQGMQLQV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTTQSLDFFPWERCEKPLTRLHVTVEGTIEENGQGMQLQV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELLISRLFTVLDHNECLIITGTEQYSEYTGIA 120  
Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELLISRLFTVLDHNECLIITGTEQYSEYTGIA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGSELMDRIYSMHI 240  
Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGSELMDRIYSMHI 916  
QY 241 FLTERKLTGVDVYKLLRLRYNEECRCNCTPGDVKLYPFYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTGVDVYKLLRLRYNEECRCNCTPGDVKLYPFYHAVESCAETADHSGQRTGT 976  
RESULT 4  
ABG72279  
ID ABG72279 standard; protein; 976 AA.  
XX ABG72279;  
AC  
XX  
DT 13-MAR-2003 (first entry)  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX  
KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective; enzyme.  
XX Homo sapiens.  
OS

XX US2002132328-A1.  
PN  
XX 19-SEP-2002.  
PD  
XX 09-OCT-2001; 2001US-00973451.  
PF  
XX 01-MAY-1998; 98US-0083768P.  
PR  
XX 30-APR-1999; 99US-00302812.  
PR  
XX (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2003-155895/15.  
XX N-PSDB; ABX14478.  
DR New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
PS Claim 28; Fig 16; 86pp; English.  
XX The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents human PARG enzyme  
XX  
SQ Sequence 976 AA;  
Query Match 100.0%; Score 1597; DB 6; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.1e-166;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFFPWERCEKPLTRLHVTVEGTIEENGQGMQLQV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTTQSLDFFPWERCEKPLTRLHVTVEGTIEENGQGMQLQV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELLISRLFTVLDHNECLIITGTEQYSEYTGIA 120  
Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELLISRLFTVLDHNECLIITGTEQYSEYTGIA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGSELMDRIYSMHI 240  
Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGSELMDRIYSMHI 916  
QY 241 FLTERKLTGVDVYKLLRLRYNEECRCNCTPGDVKLYPFYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTGVDVYKLLRLRYNEECRCNCTPGDVKLYPFYHAVESCAETADHSGQRTGT 976  
RESULT 5  
AAE25631

|    |   |                            |
|----|---|----------------------------|
| ID | AAE25631  | standard; protein; 968 AA. |
| XX |   |                            |
| AC | AAE25631;   |                            |
| XX |   |                            |
| DT | 04-NOV-2002   | (first entry)              |
| XX |   |                            |
| DE | Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).           |                            |
| XX |   |                            |
| KW | Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;   |                            |
| KW | ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;   |                            |
| KW | cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;   |                            |
| KW | Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  |                            |
| KW | cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;    |                            |
| KW | antisense therapy.  |                            |
| XX |   |                            |
| OS | Mus musculus.   |                            |
| PN | US6395543-B1.   |                            |
| XX |   |                            |
| PD | 28-MAY-2002.  |                            |
| XX |   |                            |
| PF | 23-FEB-2000; 2000US-00511507.   |                            |
| XX |   |                            |
| PR | 01-MAY-1998; 98US-0083768P.   |                            |
| PR | 30-APR-1999; 99US-00302812.   |                            |
| XX |   |                            |
| PA | (KENT ) UNIV KENTUCKY RES FOUND.  |                            |
| XX |   |                            |
| PI | Jacobson MK, Jacobson EL, Ame J, Lin W;                                   |                            |
| XX |   |                            |
| DR | WPI; 2002-535641/57.  |                            |
| DR | N-PSDB; AAD42083.   |                            |
| XX |   |                            |
| PT | New nucleic acid molecule encoding bovine poly adenosine diphosphate-     |                            |
| PT | ribose glycohydrolase involved in cellular response to DNA damage,        |                            |
| PT | inhibition of which is useful for treating neoplastic disorders and       |                            |
| PT | neurodegenerative diseases.   |                            |
| XX |   |                            |
| PS | Claim 3; Col 63-68; 77pp; English.  |                            |
| XX |   |                            |
| CC | The invention relates to an isolated nucleic acid molecule which encodes  |                            |
| CC | a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) |                            |
| CC | which catalyses release of ADP-ribose from an ADP ribose polymer. The     |                            |
| CC | invention is useful as probes and primer molecules that can be used in    |                            |
| CC | hybridisation assays and polymerase chain reaction (PCR) amplification.   |                            |
| CC | The knowledge of the nucleotide sequence of the PARG gene permits the     |                            |
| CC | preparation of antisense therapeutics containing sequences complementary  |                            |
| CC | to the mRNA of PARG gene. The antisense therapeutic are useful to treat   |                            |
| CC | neoplastic disorders and conditions caused by genotoxic oxidative stress  |                            |
| CC | e.g., cardiac disorders, neuronal disorders, reperfusion injury,          |                            |
| CC | neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  |                            |
| CC | disease. The invention is useful in gene therapy and antisense therapy.   |                            |
| XX |   |                            |
| SQ | Sequence 968 AA;  |                            |
|    | Query Match 93.7%; Score 1497; DB 5; Length 968;                          |                            |
|    | Best Local Similarity 93.7%; Pred. No. 1.3e-155;                          |                            |
|    | Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;             |                            |
| Qy | 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFPWEWCEKPLTRLHVTYEGTIEENGQMLQV 60            |                            |
| Db | 669 KTLFCYFRRVTEKKPTGLVTFTTQSLDFPWEWCEKPLTRLHVTYEGTIEENGQMLQV 728         |                            |
| Qy | 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQSYETGYA 120       |                            |
| Db | 729 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQSYETGYA 788      |                            |
| Qy | 121 ETYRWSRSHEDGSRDDCERRCTEIVADLHFRYRLDQFVPEKMRRELNKAYCGFLRPG 180         |                            |
| Db | 789 ETYRWSRSHEDGSEKDDWQRCTEIVADLHFRYRLDQFVPEKMRRELNKAYCGFLRPG 848         |                            |
| Qy | 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240       |                            |

|    |   |
|----|---|
| Db | 849 VPSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHT 908   |
| Qy | 241 FLTERKLTGVDVYKLLRLRYNEECRCNSTPGPDIKLYPFYIHAVESCAETADHSGQRTGT 300  |
| Db | 909 FLTERKLDVGKVKYKLLRLRYNEECRCNSTPGPDIKLYPFYIHAVESSAETTDMPGQKAGT 968 |

RESULT 6

AAU76022

ID AAU76022 standard; protein; 968 AA.

AC AAU76022;

XX 08-MAY-2002 (first entry)

DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;

KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;

KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;

KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;

KW neurodegenerative disease; neurological disorder; Alzheimer's disease;

XX Huntington's disease; Parkinson's disease.

OS Mus musculus.

XX

PN US6337202-B1.

XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-00511477.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

PI WPI; 2002-163240/21.

XX N-PSDB; ABK14933.

PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein

PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for

PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 2; Col 63-70; 81pp; English.

CC The present invention relates to a new poly(ADP-ribose) glycohydrolase

CC (PARG) protein which catalyses release of ADP-ribose from an ADP

CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the

CC invention is useful for generating antibodies and can be inhibited or

CC activated for diagnosing and treating neoplastic disorders such as

CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,

CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following

CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological

CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,

CC and related conditions. PARG levels may be enhanced to suppress DNA

CC repair and increase the cell's susceptibility to chemotherapy drugs.

CC Antagonists of PARG are administered to treat or prevent neoplastic

CC disorders. The present amino acid sequence represents the mouse PARG

CC protein of the invention. This protein is one of several PARG proteins

CC (AAU76020-AAU76024) of the invention

XX

SQ Sequence 968 AA;

Query Match 93.7%; Score 1497; DB 5; Length 968;

Best Local Similarity 93.7%; Pred. No. 1.3e-155;

Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Qy 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFPWEWCEKPLTRLHVTYEGTIEENGQMLQV 60

|||||

1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFPWEWCEKPLTRLHVTYEGTIEENGQMLQV 60

|||||



Db 669 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLRLHLVHYEGTIEGNRGMQLQV 728

QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTVLDHNECLITGTQEYSEYTGVA 120

Db 729 DFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTVLDHNECLITGTQEYSEYTGVA 788

QY 121 ETYRWSRSHEDGSDCERRCTEIVDAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 180

Db 789 ETYRWSRSHEDGSEKDDWQRRCTEIVDAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 848

QY 181 VSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240

Db 849 VSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHT 908

QY 241 FLTERKLTGVDVYKLLRLRYNEECNCSTPGDVKLYPFYIHAVESCAETADHSGQRTGT 300

Db 909 FLTERKLDVGKVKYKLLRLRYNEECNCSTPGDVKLYPFYIHAVESCAETDMPGQKAGT 968

RESULT 7

AAU76013

ID AAU76013 standard; protein; 968 AA.

XX AAU76013;

AC

XX

DT 08-MAY-2002 (first entry)

XX

DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX

KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;

KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;

KW inherited genetic disease; myocardial infarction; vascular stroke; aging;

KW neurodegeneration; Huntington's disease; Parkinson's disease;

KW Alzheimer's disease; neurotoxicity.

XX

OS Mus musculus.

XX

XX US6333148-B1.

PN

XX 25-DEC-2001.

PD

XX 30-APR-1999; 99US-00302812.

PF

XX 01-MAY-1998; 98US-0083768P.

PR

XX

XX (KENT ) UNIV KENTUCKY RES FOUND.

PA

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

PI

XX WPI; 2002-153820/20.

DR

XX N-PSDB; ABK14495.

DR

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,

PT useful potentially for treating diseases associated with DNA damage, e.g.

PT cancer.

PT

XX Claim 3; Col 63-68; 80pp; English.

PS

XX

CC The present invention relates to a new method for screening compounds for

CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine

CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are

CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and

CC are used to treat or prevent any condition associated with DNA damage

CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,

CC vascular stroke, aging and neurodegeneration e.g. Huntington's,

CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.

CC Compounds identified by the new method are more effective than known

CC inhibitors and have fewer side effects. The present amino acid sequence

CC represents the mouse PARG protein of the invention. This protein is one

CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the

CC invention

XX

SQ Sequence 968 AA;

Query Match 93.7%; Score 1497; DB 5; Length 968;

Best Local Similarity 93.7%; Pred. No. 1.3e-155;

Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLRLHLVHYEGTIEGNRGMQLQV 60

Db 669 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLRLHLVHYEGTIEGNRGMQLQV 728

QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTVLDHNECLITGTQEYSEYTGVA 120

Db 729 DFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTVLDHNECLITGTQEYSEYTGVA 788

QY 121 ETYRWSRSHEDGSDCERRCTEIVDAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 180

Db 789 ETYRWSRSHEDGSEKDDWQRRCTEIVDAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 848

QY 181 VSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240

Db 849 VSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHT 908

QY 241 FLTERKLTGVDVYKLLRLRYNEECNCSTPGDVKLYPFYIHAVESCAETADHSGQRTGT 300

Db 909 FLTERKLDVGKVKYKLLRLRYNEECNCSTPGDVKLYPFYIHAVESCAETDMPGQKAGT 968

RESULT 8

ABG72280

ID ABG72280 standard; protein; 968 AA.

XX

AC ABG72280;

XX

DT 13-MAR-2003 (first entry)

XX

DE Murine poly(ADP-ribose) glycohydrolase (PARG) enzyme.

XX

KW Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;

KW cellular response; DNA damage; neoplastic disorder inducing agent;

KW oxidative stress; neoplastic disorder; myocardial infarction;

KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;

KW Parkinson's disease; Huntington's disease; inborn genetic error;

KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;

KW cytosolic; neuroprotective; neurotropic; antiparkinsonian; cardiant;

KW vasotropic; anticonvulsant; cerebroprotective; enzyme.

XX

OS Mus musculus.

XX

XX US2002132328-A1.

PN

XX 19-SEP-2002.

PD

XX 09-OCT-2001; 2001US-00973451.

PF

XX 01-MAY-1998; 98US-0083768P.

PR

XX 30-APR-1999; 99US-00302812.

PR

XX

XX (JACO/) JACOBSON M K.

PA (JACO/) JACOBSON E L.

PA (AMEJ/) AME J.

PA (LINW/) LIN W.

XX

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2003-155895/15.

DR N-PSDB; ABX14479.

DR

XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)

PT glycohydrolase activity, for preventing, treating, or ameliorating a

PT disease condition, e.g. neoplastic disorder, myocardial infarction or

PT vascular stroke.

PT

XX Claim 28; Fig 16; 86pp; English.

PS

XX

The present invention relates to the isolation of poly (ADP-ribose) glycohydrolase (PARG) from several species, and the polynucleotide sequences encoding them. Methods for inhibiting PARG expression or overexpressing PARG are also disclosed. PARG is involved in the cellular response to DNA damage, and is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. The polynucleotide sequences encoding PARG and PARG modulators are useful for preventing, treating, or ameliorating diseases such as neoplastic disorders, myocardial infarction, vascular stroke, neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's disease), inborn genetic errors, reperfusion following ischaemia, aging, and neurotoxicity. The polynucleotide sequences are also useful in gene therapy. The methods are useful for identifying an agent that modulates PARG activity, identifying a mutant PARG allele in an individual, or screening candidate molecules for PARG modulating activity. The present sequence represents murine PARG enzyme

Sequence 968 AA;

XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-535641/57.  
XX N-PSDB; AAD42081.  
DR  
DR  
XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
XX Claim 3; Col 47-45; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is bPARG



Db 678 KTLFCYFRRVTEKPTGLVTFTTRQSLDFPEWERCXKLLTRLHVYEGTIEGNGQMLQV 737

QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIIITGTEQYSEYTGIA 120

Db 738 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIIITGTEQYSEYTGIA 797

QY 121 ETVRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180

Db 798 ETVRWSRSHEDRSDQRTTEIVADALHFRYLDQFVPEKIRRELKAYCGFLRPG 857

QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDIYSMHI 240

Db 858 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDIYSMHT 917

QY 241 FLTERKLTVDVYKLLRLRYNEECNCSTPGPDIKLYPFYHVESCAETADHSGQRTG 299

Db 918 FLTERKLTGVEVYKLLRLRYNEECNCSTPGPDIKLYPFYHVESCTQTNTNPGQRTG 976

RESULT 12

ABG72278

ID ABG72278 standard; protein; 977 AA.

XX

AC ABG72278;

XX

DT 13-MAR-2003 (first entry)

XX

DE Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.

XX

KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression; cellular response; DNA damage; neoplastic disorder inducing agent; oxidative stress; neoplastic disorder; myocardial infarction; vascular stroke; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; Huntington's disease; inborn genetic error; reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic; neuroprotective; neurotropic; antiparkinsonian; cardiant; vasotropic; anticonvulsant; cerebroprotective; enzyme.

XX

OS Bos taurus.

XX

PN US2002132328-A1.

XX

PD 19-SEP-2002.

XX

PF 09-OCT-2001; 2001US-00973451.

XX

PR 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

XX

PA (JACO/) JACOBSON M K.

PA (JACO/) JACOBSON E L.

PA (AMEJ/) AME J.

PA (LINW/) LIN W.

XX

PI Jacobson MK, Jacobson EL, Ame J, Lin W;

XX

DR WPI; 2003-155895/15.

DR N-PSDB; ABX14477.

XX

PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose) glycohydrolase activity, for preventing, treating, or ameliorating a disease condition, e.g. neoplastic disorder, myocardial infarction or vascular stroke.

XX

PS Claim 28; Fig 16; 86pp; English.

XX

CC The present invention relates to the isolation of poly(ADP-ribose) glycohydrolase (PARG) from several species, and the polynucleotide sequences encoding them. Methods for inhibiting PARG expression or overexpressing PARG are also disclosed. PARG is involved in the cellular response to DNA damage, and is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. The polynucleotide sequences encoding PARG and PARG modulators are useful for

CC preventing, treating, or ameliorating diseases such as neoplastic disorders, myocardial infarction, vascular stroke, neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's disease), inborn genetic errors, reperfusion following ischaemia, aging, and neurotoxicity. The polynucleotide sequences are also useful in gene therapy. The methods are useful for identifying an agent that modulates PARG activity, identifying a mutant PARG allele in an individual, or screening candidate molecules for PARG modulating activity. The present sequence represents bovine PARG enzyme

XX

SQ Sequence 977 AA;

Query Match 93.5%; Score 1493; DB 6; Length 977;

Best Local Similarity 94.0%; Pred. No. 3.6e-155;

Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKPTGLVTFTTRQSLDFPEWERCXKLLTRLHVYEGTIEGNGQMLQV 60

Db 678 KTLFCYFRRVTEKPTGLVTFTTRQSLDFPEWERCXKLLTRLHVYEGTIEGNGQMLQV 737

QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIIITGTEQYSEYTGIA 120

Db 738 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIIITGTEQYSEYTGIA 797

QY 121 ETVRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180

Db 798 ETVRWSRSHEDRSDQRTTEIVADALHFRYLDQFVPEKIRRELKAYCGFLRPG 857

QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDIYSMHI 240

Db 858 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDIYSMHT 917

QY 241 FLTERKLTVDVYKLLRLRYNEECNCSTPGPDIKLYPFYHVESCAETADHSGQRTG 299

Db 918 FLTERKLTGVEVYKLLRLRYNEECNCSTPGPDIKLYPFYHVESCTQTNTNPGQRTG 976

RESULT 13

ABB59491

ID ABB59491 standard; protein; 768 AA.

XX

AC ABB59491;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 5265.

XX

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL03594.

XX

CC New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

PT

PT

XX

PS Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.



XX CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 768 AA;  
Query Match 38.8%; Score 619.5; DB 4; Length 768;  
Best Local Similarity 46.4%; Pred. No. 9e-59;  
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;  
QY 1 KTLFCYFRRV--TEKK-----PTGLVTFTRQS--LEDFPWERCEKPL--TRLHVTVEGTI 50  
Db 297 KCTMHYFRRVCPTERDASNVPTGVVTFVRRSGLPEHLIDWSQAAPLGDVPLHVDAGETI 356  
QY 51 EENGQGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTTEVLDHNECLIITGT 110  
Db 357 EDEGIGLLQVDFANKYLGGLGGVLCGVCQEEIRFVICPELLVKGKLFTECLRPFEALVMLGA 416  
QY 111 EQYSEYTGAEYRWSRSHEDGSRDDCERRCTEIVADALHFRRYLDQFVPEKMRRELN 170  
Db 417 ERYSNYTGAGSFEWSGNFEDSTPRDSSGRRQTAIVAIDALHFAQSHHQYREDLMERELN 476  
QY 171 KAYCGFLR-----PGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAEERDVVYF 224  
Db 477 KAYIGFVHMVTPPPG-----VATGNWCGGAFGGDSYLLKALLQMLMVCALGRPLAYY 528  
QY 225 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECNCSTPGP-----DIKLYP 278  
Db 529 TFGNVFRRDDFHEMWLLFRNDGTTVQQLWS--ILRSYRLIKESSKEPRENKASKKLYD 587  
QY 279 FI 280  
Db 588 FI 589  
RESULT 14  
AAE25632  
ID AAE25632 standard; protein; 768 AA.  
XX AC AAE25632;  
XX 04-NOV-2002 (first entry)  
XX Fruit fly poly adenosine diphosphate-ribose glycohydrolase (PARG).  
XX Poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy; fruit fly.  
XX OS Drosophila melanogaster.  
XX PN US6395543-B1.  
XX 28-MAY-2002.  
XX 23-FEB-2000; 2000US-00511507.  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX

PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
DR N-PSDB; AAD42084.  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX Example 13; Col 71-76; 77pp; English.  
XX The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is fruit fly PARG  
XX SQ Sequence 768 AA;  
Query Match 38.8%; Score 619.5; DB 5; Length 768;  
Best Local Similarity 46.4%; Pred. No. 9e-59;  
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;  
QY 1 KTLFCYFRRV--TEKK-----PTGLVTFTRQS--LEDFPWERCEKPL--TRLHVTVEGTI 50  
Db 297 KCTMHYFRRVCPTERDASNVPTGVVTFVRRSGLPEHLIDWSQAAPLGDVPLHVDAGETI 356  
QY 51 EENGQGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTTEVLDHNECLIITGT 110  
Db 357 EDEGIGLLQVDFANKYLGGLGGVLCGVCQEEIRFVICPELLVKGKLFTECLRPFEALVMLGA 416  
QY 111 EQYSEYTGAEYRWSRSHEDGSRDDCERRCTEIVADALHFRRYLDQFVPEKMRRELN 170  
Db 417 ERYSNYTGAGSFEWSGNFEDSTPRDSSGRRQTAIVAIDALHFAQSHHQYREDLMERELN 476  
QY 171 KAYCGFLR-----PGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAEERDVVYF 224  
Db 477 KAYIGFVHMVTPPPG-----VATGNWCGGAFGGDSYLLKALLQMLMVCALGRPLAYY 528  
QY 225 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECNCSTPGP-----DIKLYP 278  
Db 529 TFGNVFRRDDFHEMWLLFRNDGTTVQQLWS--ILRSYRLIKESSKEPRENKASKKLYD 587  
QY 279 FI 280  
Db 588 FI 589  
RESULT 15  
AAU76023  
ID AAU76023 standard; protein; 768 AA.  
XX AC AAU76023;  
XX 08-MAY-2002 (first entry)  
XX Fruit fly poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX Fruit fly; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 9.61306 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977  
Perfect score: 1597  
Sequence: 1 KTLFCYFRRVTEKKPTGLVT.....YHAVESCAETADHSGQRTGT 300

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID                 | Description       |
|------------|-------|-------------|--------|-----------------------|-------------------|
| 1          | 1597  | 100.0       | 976    | 4 US-09-302-812-4     | Sequence 4, Appli |
| 2          | 1597  | 100.0       | 976    | 4 US-09-511-477-4     | Sequence 4, Appli |
| 3          | 1597  | 100.0       | 976    | 4 US-09-511-507-4     | Sequence 4, Appli |
| 4          | 1497  | 93.7        | 968    | 4 US-09-302-812-6     | Sequence 6, Appli |
| 5          | 1497  | 93.7        | 968    | 4 US-09-511-477-6     | Sequence 6, Appli |
| 6          | 1497  | 93.7        | 968    | 4 US-09-511-507-6     | Sequence 6, Appli |
| 7          | 1493  | 93.5        | 977    | 4 US-09-302-812-2     | Sequence 2, Appli |
| 8          | 1493  | 93.5        | 977    | 4 US-09-511-477-2     | Sequence 2, Appli |
| 9          | 1493  | 93.5        | 977    | 4 US-09-511-507-2     | Sequence 2, Appli |
| 10         | 619.5 | 38.8        | 768    | 4 US-09-302-812-8     | Sequence 8, Appli |
| 11         | 619.5 | 38.8        | 768    | 4 US-09-511-477-8     | Sequence 8, Appli |
| 12         | 619.5 | 38.8        | 768    | 4 US-09-511-507-8     | Sequence 8, Appli |
| 13         | 371.5 | 23.3        | 726    | 4 US-09-302-812-10    | Sequence 10, Appl |
| 14         | 371.5 | 23.3        | 726    | 4 US-09-511-477-10    | Sequence 10, Appl |
| 15         | 371.5 | 23.3        | 726    | 4 US-09-511-507-10    | Sequence 10, Appl |
| 16         | 167   | 10.5        | 31     | 4 US-09-302-812-11    | Sequence 11, Appl |
| 17         | 167   | 10.5        | 31     | 4 US-09-511-477-11    | Sequence 11, Appl |
| 18         | 167   | 10.5        | 31     | 4 US-09-511-507-11    | Sequence 11, Appl |
| 19         | 149   | 9.3         | 29     | 4 US-09-302-812-12    | Sequence 12, Appl |
| 20         | 149   | 9.3         | 29     | 4 US-09-511-477-12    | Sequence 12, Appl |
| 21         | 149   | 9.3         | 29     | 4 US-09-511-507-12    | Sequence 12, Appl |
| 22         | 91.5  | 5.7         | 261    | 2 US-08-691-814B-4    | Sequence 4, Appli |
| 23         | 90    | 5.6         | 749    | 4 US-09-328-352-7588  | Sequence 7588, Ap |
| 24         | 89    | 5.6         | 205    | 4 US-08-630-915A-16   | Sequence 16, Appl |
| 25         | 85    | 5.3         | 355    | 4 US-09-489-039A-8367 | Sequence 8367, Ap |
| 26         | 85    | 5.3         | 580    | 3 US-09-234-393-48    | Sequence 48, Appl |
| 27         | 85    | 5.3         | 580    | 4 US-09-865-171-48    | Sequence 48, Appl |

|    |      |     |     |                      |                   |
|----|------|-----|-----|----------------------|-------------------|
| 28 | 85   | 5.3 | 581 | 3 US-09-234-393-20   | Sequence 20, Appl |
| 29 | 85   | 5.3 | 581 | 3 US-09-234-393-44   | Sequence 44, Appl |
| 30 | 85   | 5.3 | 581 | 3 US-09-234-393-46   | Sequence 46, Appl |
| 31 | 85   | 5.3 | 581 | 4 US-09-865-171-20   | Sequence 20, Appl |
| 32 | 85   | 5.3 | 581 | 4 US-09-865-171-44   | Sequence 44, Appl |
| 33 | 85   | 5.3 | 581 | 4 US-09-865-171-46   | Sequence 46, Appl |
| 34 | 85   | 5.3 | 581 | 4 US-09-398-395A-48  | Sequence 48, Appl |
| 35 | 85   | 5.3 | 581 | 4 US-09-887-586A-48  | Sequence 48, Appl |
| 36 | 85   | 5.3 | 581 | 4 US-09-895-752-48   | Sequence 48, Appl |
| 37 | 85   | 5.3 | 581 | 4 US-09-903-012B-48  | Sequence 48, Appl |
| 38 | 85   | 5.3 | 581 | 4 US-09-900-797-48   | Sequence 48, Appl |
| 39 | 83.5 | 5.2 | 505 | 4 US-09-620-405B-478 | Sequence 478, App |
| 40 | 83.5 | 5.2 | 505 | 4 US-09-620-405B-485 | Sequence 485, App |
| 41 | 83.5 | 5.2 | 505 | 4 US-09-604-287A-478 | Sequence 478, App |
| 42 | 83.5 | 5.2 | 505 | 4 US-09-604-287A-485 | Sequence 485, App |
| 43 | 83.5 | 5.2 | 505 | 4 US-09-834-759-478  | Sequence 478, App |
| 44 | 83.5 | 5.2 | 505 | 4 US-09-834-759-485  | Sequence 485, App |
| 45 | 83.5 | 5.2 | 915 | 2 US-08-480-917-2    | Sequence 2, Appli |

ALIGNMENTS

RESULT 1  
US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA2 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

|                       |        |   |          |            |   |        |     |
|-----------------------|--------|---|----------|------------|---|--------|-----|
| Query Match           | 100.0% | Score   | 1597     | DB         | 4 | Length | 976 |
| Best Local Similarity | 100.0% | Pred. No.   | 2.1e-171 |            |   |        |     |
| Matches               | 300    | Conservative  | 0        | Mismatches | 0 | Indels | 0   |
|                       |        | Gaps  | 0        |            |   |        |     |
| Qy                    | 1      | KTLFCYFRRVTEKKPTGLVTFTTROSLEDFPEWERCCKPLTRLHVTYEGTIEENGQGMLOV | 60       |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Db                    | 677    | KTLFCYFRRVTEKKPTGLVTFTTROSLEDFPEWERCCKPLTRLHVTYEGTIEENGQGMLOV | 736      |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Qy                    | 61     | DFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTEVLHDHNECLITGTQEYSEYTGVA   | 120      |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Db                    | 737    | DFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTEVLHDHNECLITGTQEYSEYTGVA   | 796      |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Qy                    | 121    | ETVRSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG    | 180      |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Db                    | 797    | ETVRSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG    | 856      |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Qy                    | 181    | VSENL SAVATGNWCGAFGGDARLKALIQILAAAAAERD VVYFTFGDSELMRDIYSMHI  | 240      |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Db                    | 857    | VSENL SAVATGNWCGAFGGDARLKALIQILAAAAAERD VVYFTFGDSELMRDIYSMHI  | 916      |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Qy                    | 241    | FLTERKLTVDVYKLLRLRYNEECRN CSTPGPDIKLYPFYIHAVESCAETADHSGQRTGT  | 300      |            |   |        |     |
|                       |        |   |          |            |   |        |     |
| Db                    | 917    | FLTERKLTVDVYKLLRLRYNEECRN CSTPGPDIKLYPFYIHAVESCAETADHSGQRTGT  | 976      |            |   |        |     |
|                       |        |   |          |            |   |        |     |

RESULT 2  
US-09-511-477-4  
; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-477-4  
Query Match 100.0%; Score 1597; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 2.1e-171;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGYA 120  
Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGYA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 916  
QY 241 FLTERKLTGVDVYKLLRLRYNEECRCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTGVDVYKLLRLRYNEECRCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976  
RESULT 3  
US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:

US-09-511-507-4  
Query Match 100.0%; Score 1597; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 2.1e-171;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGYA 120  
Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGYA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 916  
QY 241 FLTERKLTGVDVYKLLRLRYNEECRCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTGVDVYKLLRLRYNEECRCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976  
RESULT 4  
US-09-302-812-6  
; Sequence 6, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-302-812-6  
Query Match 93.7%; Score 1497; DB 4; Length 968;  
Best Local Similarity 93.7%; Pred. No. 4.3e-160;  
Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;  
QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 60  
Db 669 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 728  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGYA 120  
Db 729 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGYA 788  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 789 ETYRWSRSHEDGSEKDDQRRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 848  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 849 VPSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHT 908  
QY 241 FLTERKLTGVDVYKLLRLRYNEECRCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300



```
Db 909 FLTERKLDVGKVKYKLLRYNEECRNCSTPGPDIKLYPFIYHAVESSAETDMPGQKAGT 968
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: ||
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-477-6

RESULT 5
US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAID 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-477-6

Query Match 93.7%; Score 1497; DB 4; Length 968;
Best Local Similarity 93.7%; Pred. No. 4.3e-160;
Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCCKPLTRLHVITYEGTIEENGQGMLOV 60
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 669 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCCKPLTRLHVITYEGTIEENGQGMLOV 728
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGYA 120
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 729 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGYA 788
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVDAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 180
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 789 ETYRWSRSHEDGSRDDCERRCTEIVDAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 848
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 849 VPSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHT 908
QY 241 FLTERKLTGVDVYKLLRYNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 909 FLTERKLDVGKVKYKLLRYNEECRNCSTPGPDIKLYPFIYHAVESSAETDMPGQKAGT 968

RESULT 6
US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAID 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
```

```
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-507-6

Query Match 93.7%; Score 1497; DB 4; Length 968;
Best Local Similarity 93.7%; Pred. No. 4.3e-160;
Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCCKPLTRLHVITYEGTIEENGQGMLOV 60
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 669 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCCKPLTRLHVITYEGTIEENGQGMLOV 728
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGYA 120
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 729 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGYA 788
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVDAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 180
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 789 ETYRWSRSHEDGSRDDCERRCTEIVDAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 848
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 849 VPSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHT 908
QY 241 FLTERKLTGVDVYKLLRYNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 909 FLTERKLDVGKVKYKLLRYNEECRNCSTPGPDIKLYPFIYHAVESSAETDMPGQKAGT 968

RESULT 7
US-09-302-812-2
; Sequence 2, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAID 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-302-812-2

Query Match 93.5%; Score 1493; DB 4; Length 977;
Best Local Similarity 94.0%; Pred. No. 1.2e-159;
Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCCKPLTRLHVITYEGTIEENGQGMLOV 60
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 678 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCCKPLTRLHVITYEGTIEENGQGMLOV 737
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGYA 120
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 738 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGYA 797
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVDAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 180
||||| ||| ||||||||||||||||||||||||||||||||||||||| ||| ||| |||: |||
Db 798 ETYRWSRSHEDGSRDDCERRCTEIVDAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 857
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240
```

Db 858 VSSNL SAVATGNWCGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMRDIYSMHT 917  
QY 241 FLTERKLTVDVYKLLRLRYNEECNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTG 299  
Db 918 FLTERKLTVDVYKLLRLRYNEECNCSTPGPDIKLYPFIYHAVESCTQTNTNPGQRTG 976  
RESULT 8  
US-09-511-477-2  
; Sequence 2, Application US/095111477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-2

Query Match 93.5%; Score 1493; DB 4; Length 977;  
Best Local Similarity 94.0%; Pred. No. 1.2e-159;  
Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;  
QY 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFFPEWERCCKPLTRLHLVHYEGTIEENGQGMQV 60  
Db 678 KTLFCYFRRVTEKKPTGLVTFTTQSLDFFPEWERCCKPLTRLHLVHYEGTIEENGQGMQV 737  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGVA 120  
Db 738 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGVA 797  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 798 ETYRWSRSHEDRSDWQRTTEIVADALHFRYLDQFVPEKIRRELKAYCGFLRPG 857  
QY 181 VSSNL SAVATGNWCGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMRDIYSMHI 240  
Db 858 VSSNL SAVATGNWCGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMRDIYSMHT 917  
QY 241 FLTERKLTVDVYKLLRLRYNEECNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTG 299  
Db 918 FLTERKLTVDVYKLLRLRYNEECNCSTPGPDIKLYPFIYHAVESCTQTNTNPGQRTG 976  
RESULT 9  
US-09-511-507-2  
; Sequence 2, Application US/095111507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23

; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-2  
Query Match 93.5%; Score 1493; DB 4; Length 977;  
Best Local Similarity 94.0%; Pred. No. 1.2e-159;  
Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;  
QY 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDFFPEWERCCKPLTRLHLVHYEGTIEENGQGMQV 60  
Db 678 KTLFCYFRRVTEKKPTGLVTFTTQSLDFFPEWERCCKPLTRLHLVHYEGTIEENGQGMQV 737  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGVA 120  
Db 738 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGVA 797  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 798 ETYRWSRSHEDRSDWQRTTEIVADALHFRYLDQFVPEKIRRELKAYCGFLRPG 857  
QY 181 VSSNL SAVATGNWCGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMRDIYSMHI 240  
Db 858 VSSNL SAVATGNWCGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMRDIYSMHT 917  
QY 241 FLTERKLTVDVYKLLRLRYNEECNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTG 299  
Db 918 FLTERKLTVDVYKLLRLRYNEECNCSTPGPDIKLYPFIYHAVESCTQTNTNPGQRTG 976

RESULT 10  
US-09-302-812-8  
; Sequence 8, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-302-812-8

Query Match 38.8%; Score 619.5; DB 4; Length 768;  
Best Local Similarity 46.4%; Pred. No. 6.3e-61;  
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;  
QY 1 KTLFCYFRRV--TEKK----PTGLVTFTTQSLDFFPEWERCCKPL--TRLHYVEGTI 50  
Db 297 KCIIMHYFRRVCPTEERDASNVPTGVVTFVRRSGLPEHLIDWSQSAAPLGDVPLHVDAGTI 356  
QY 51 EENGQGMQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGT 110  
Db 357 EDEGIGLLQVDFANKYLGGLGVGHGCVQEEIRFVICPELLVGLKFTCLRPFEALVMLGA 416  
QY 111 EQYSEYTGAEITYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELN 170

```
Db 417 ERYSNYTGAGSFEWSGNFEDSTPRDSSGRRQTAVAIADALHFAQSHHQYREDLIMERELN 476
QY 171 KAYCGFLR-----PGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVVF 224
Db 477 KAYIGFVHMVTPPPG-----VATGNWCGGAFGGDSYLLKALLQMLVCAQLGRPLAYY 528
QY 225 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECNCSTPGP-----DIKLYP 278
Db 529 TFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRSYSRLIKESSKEPRENKASKKLYD 587
QY 279 FI 280
Db 588 FI 589

RESULT 11
US-09-511-477-8
; Sequence 8, Application US/095111477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-477-8

Query Match 38.8%; Score 619.5; DB 4; Length 768;
Best Local Similarity 46.4%; Pred. No. 6.3e-61;
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;

QY 1 KTLFCYFRRV--TEKK-----PTGLVTFTQSQ--LEDFFPWERCEKPL--TRLHVTVYEGTI 50
Db 297 KCIIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPEHLIDWSQSAAPLGDVPLHVDAEGTI 356
QY 51 EENGQGMLOQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTTEVLHDNECLIITGT 110
Db 357 EDEGIGLLQVDFANKYLGGLGVGHGCVQEEIRFVICPELLVGVKLFTECLRPFEALVMLGA 416
QY 111 EQYSEYTGAEYRWRSRSHEDGSDERDDCERRCTEIVAIDALHFRRLYLDQFVPEKMRRELN 170
Db 417 ERYSNYTGAGSFEWSGNFEDSTPRDSSGRRQTAVAIADALHFAQSHHQYREDLIMERELN 476
QY 171 KAYCGFLR-----PGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVVF 224
Db 477 KAYIGFVHMVTPPPG-----VATGNWCGGAFGGDSYLLKALLQMLVCAQLGRPLAYY 528
QY 225 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECNCSTPGP-----DIKLYP 278
Db 529 TFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRSYSRLIKESSKEPRENKASKKLYD 587
QY 279 FI 280
Db 588 FI 589

RESULT 12
US-09-511-507-8
; Sequence 8, Application US/095111507
```

```
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

Query Match 38.8%; Score 619.5; DB 4; Length 768;
Best Local Similarity 46.4%; Pred. No. 6.3e-61;
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;

QY 1 KTLFCYFRRV--TEKK-----PTGLVTFTQSQ--LEDFFPWERCEKPL--TRLHVTVYEGTI 50
Db 297 KCIIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPEHLIDWSQSAAPLGDVPLHVDAEGTI 356
QY 51 EENGQGMLOQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTTEVLHDNECLIITGT 110
Db 357 EDEGIGLLQVDFANKYLGGLGVGHGCVQEEIRFVICPELLVGVKLFTECLRPFEALVMLGA 416
QY 111 EQYSEYTGAEYRWRSRSHEDGSDERDDCERRCTEIVAIDALHFRRLYLDQFVPEKMRRELN 170
Db 417 ERYSNYTGAGSFEWSGNFEDSTPRDSSGRRQTAVAIADALHFAQSHHQYREDLIMERELN 476
QY 171 KAYCGFLR-----PGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVVF 224
Db 477 KAYIGFVHMVTPPPG-----VATGNWCGGAFGGDSYLLKALLQMLVCAQLGRPLAYY 528
QY 225 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECNCSTPGP-----DIKLYP 278
Db 529 TFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRSYSRLIKESSKEPRENKASKKLYD 587
QY 279 FI 280
Db 588 FI 589

RESULT 13
US-09-302-812-10
; Sequence 10, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
```

```

; FEATURE:
US-09-302-812-10
    Query Match      23.3%; Score 371.5; DB 4; Length 726;
    Best Local Similarity 35.3%; Pred. No. 6.8e-33;
    Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRRVTEKKPTGLVTF--TRQSLDF-PEWERCCKPLTRLHVT--YEGTIEENGQ 55
Db 456 KFLFTYFDKMSMDPPDGAVSFLTKMDKDTFNEEWK--DKKLRSLEPEVEFFDEMLIEDTA 513

QY 56 GMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSE 115
Db 514 LCTQVDFANEHLGGVNLHGVSQVEEIRFLMCPMMVGMVLLCEKMKQLEAISIVGAYVFSS 573

QY 116 YTGVAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFRR-----YLDQFVPEK 164
Db 574 YTGYGHTLKWAELOPNHRSRQNTNEFRDRFGLRVETIAIDAILFKGSKLDCQTEQLNKAN 633

QY 165 MRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYF 224
Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGGAFNGDKPLKFIQIVIAAGVADRPLHFC 692

QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250
Db 693 SFGEPELAACKKIIERMKQKDVTLG 718

RESULT 14
US-09-511-477-10
; Sequence 10, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-477-10
    Query Match      23.3%; Score 371.5; DB 4; Length 726;
    Best Local Similarity 35.3%; Pred. No. 6.8e-33;
    Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRRVTEKKPTGLVTF--TRQSLDF-PEWERCCKPLTRLHVT--YEGTIEENGQ 55
Db 456 KFLFTYFDKMSMDPPDGAVSFLTKMDKDTFNEEWK--DKKLRSLEPEVEFFDEMLIEDTA 513

QY 56 GMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSE 115
Db 514 LCTQVDFANEHLGGVNLHGVSQVEEIRFLMCPMMVGMVLLCEKMKQLEAISIVGAYVFSS 573

QY 116 YTGVAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFRR-----YLDQFVPEK 164
Db 574 YTGYGHTLKWAELOPNHRSRQNTNEFRDRFGLRVETIAIDAILFKGSKLDCQTEQLNKAN 633

QY 165 MRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYF 224
Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGGAFNGDKPLKFIQIVIAAGVADRPLHFC 692

QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250
Db 693 SFGEPELAACKKIIERMKQKDVTLG 718

RESULT 14
US-09-511-477-10
; Sequence 10, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-477-10
    Query Match      23.3%; Score 371.5; DB 4; Length 726;
    Best Local Similarity 35.3%; Pred. No. 6.8e-33;
    Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRRVTEKKPTGLVTF--TRQSLDF-PEWERCCKPLTRLHVT--YEGTIEENGQ 55
Db 456 KFLFTYFDKMSMDPPDGAVSFLTKMDKDTFNEEWK--DKKLRSLEPEVEFFDEMLIEDTA 513

QY 56 GMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSE 115
Db 514 LCTQVDFANEHLGGVNLHGVSQVEEIRFLMCPMMVGMVLLCEKMKQLEAISIVGAYVFSS 573

QY 116 YTGVAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFRR-----YLDQFVPEK 164
Db 574 YTGYGHTLKWAELOPNHRSRQNTNEFRDRFGLRVETIAIDAILFKGSKLDCQTEQLNKAN 633

QY 165 MRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYF 224
Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGGAFNGDKPLKFIQIVIAAGVADRPLHFC 692
```

```

QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250
Db 693 SFGEPELAACKKIIERMKQKDVTLG 718

RESULT 15
US-09-511-507-10
; Sequence 10, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-507-10
    Query Match      23.3%; Score 371.5; DB 4; Length 726;
    Best Local Similarity 35.3%; Pred. No. 6.8e-33;
    Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRRVTEKKPTGLVTF--TRQSLDF-PEWERCCKPLTRLHVT--YEGTIEENGQ 55
Db 456 KFLFTYFDKMSMDPPDGAVSFLTKMDKDTFNEEWK--DKKLRSLEPEVEFFDEMLIEDTA 513

QY 56 GMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSE 115
Db 514 LCTQVDFANEHLGGVNLHGVSQVEEIRFLMCPMMVGMVLLCEKMKQLEAISIVGAYVFSS 573

QY 116 YTGVAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFRR-----YLDQFVPEK 164
Db 574 YTGYGHTLKWAELOPNHRSRQNTNEFRDRFGLRVETIAIDAILFKGSKLDCQTEQLNKAN 633

QY 165 MRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYF 224
Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGGAFNGDKPLKFIQIVIAAGVADRPLHFC 692

QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250
Db 693 SFGEPELAACKKIIERMKQKDVTLG 718

Search completed: May 26, 2004, 18:49:26
Job time : 10.6131 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 25.0302 seconds  
(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977  
Perfect score: 1597  
Sequence: 1 KTLFCYFRRVTEKKPTGLVT.....YHAVESCAETADHSGQRTGT 300

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

|     |  |
|-----|--|
| 1:  | /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*  |
| 2:  | /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*   |
| 3:  | /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*  |
| 4:  | /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*  |
| 5:  | /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*  |
| 6:  | /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.* |
| 7:  | /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*  |
| 8:  | /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*  |
| 9:  | /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.* |
| 10: | /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.* |
| 11: | /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.* |
| 12: | /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*  |
| 13: | /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.* |
| 14: | /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.* |
| 15: | /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.* |
| 16: | /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*  |
| 17: | /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*  |
| 18: | /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description          |
|------------|-------|---------------|--------|-------|----------------------|
| 1          | 1597  | 100.0         | 976    | 9     | US-09-973-451-4      |
| 2          | 1497  | 93.7          | 968    | 9     | US-09-973-451-6      |
| 3          | 1493  | 93.5          | 977    | 9     | US-09-973-451-2      |
| 4          | 619.5 | 38.8          | 768    | 9     | US-09-973-451-8      |
| 5          | 489   | 30.6          | 546    | 12    | US-10-425-114-60000  |
| 6          | 371.5 | 23.3          | 726    | 9     | US-09-973-451-10     |
| 7          | 191.5 | 12.0          | 200    | 12    | US-10-424-599-184988 |
| 8          | 188   | 11.8          | 180    | 12    | US-10-424-599-156445 |
| 9          | 167   | 10.5          | 31     | 9     | US-09-973-451-11     |
| 10         | 149   | 9.3           | 29     | 9     | US-09-973-451-12     |
| 11         | 99.5  | 6.2           | 300    | 12    | US-10-424-599-233915 |
| 12         | 97.5  | 6.1           | 395    | 12    | US-10-282-122A-51256 |
| 13         | 94.5  | 5.9           | 730    | 15    | US-10-369-493-19591  |
| 14         | 92.5  | 5.8           | 7349   | 14    | US-10-314-657-46     |
| 15         | 91.5  | 5.7           | 261    | 14    | US-10-247-671-134    |

|    |      |     |     |    |                      |                   |
|----|------|-----|-----|----|----------------------|-------------------|
| 16 | 91.5 | 5.7 | 261 | 15 | US-10-435-696-27     | Sequence 27, Appl |
| 17 | 89   | 5.6 | 205 | 9  | US-09-879-957-16     | Sequence 16, Appl |
| 18 | 88   | 5.5 | 442 | 15 | US-10-369-493-11707  | Sequence 11707, A |
| 19 | 85   | 5.3 | 581 | 9  | US-09-887-586A-48    | Sequence 48, Appl |
| 20 | 85   | 5.3 | 581 | 9  | US-09-903-012-48     | Sequence 48, Appl |
| 21 | 85   | 5.3 | 581 | 10 | US-09-900-797-48     | Sequence 48, Appl |
| 22 | 85   | 5.3 | 581 | 12 | US-09-893-820-48     | Sequence 48, Appl |
| 23 | 85   | 5.3 | 581 | 13 | US-10-041-007-18     | Sequence 18, Appl |
| 24 | 84.5 | 5.3 | 982 | 14 | US-10-156-761-11410  | Sequence 11410, A |
| 25 | 83.5 | 5.2 | 153 | 12 | US-10-282-122A-56462 | Sequence 56462, A |
| 26 | 83.5 | 5.2 | 505 | 9  | US-09-604-287A-478   | Sequence 478, App |
| 27 | 83.5 | 5.2 | 505 | 9  | US-09-604-287A-485   | Sequence 485, App |
| 28 | 83.5 | 5.2 | 505 | 10 | US-09-551-621-478    | Sequence 478, App |
| 29 | 83.5 | 5.2 | 505 | 13 | US-10-007-805-478    | Sequence 478, App |
| 30 | 83.5 | 5.2 | 505 | 13 | US-10-007-805-485    | Sequence 485, App |
| 31 | 83.5 | 5.2 | 505 | 14 | US-10-076-622-478    | Sequence 478, App |
| 32 | 83.5 | 5.2 | 505 | 14 | US-10-076-622-485    | Sequence 485, App |
| 33 | 83.5 | 5.2 | 505 | 14 | US-10-124-805-478    | Sequence 478, App |
| 34 | 83.5 | 5.2 | 505 | 14 | US-10-124-805-485    | Sequence 485, App |
| 35 | 82.5 | 5.2 | 294 | 12 | US-10-424-599-249573 | Sequence 249573,  |
| 36 | 82.5 | 5.2 | 452 | 12 | US-10-425-114-41682  | Sequence 41682, A |
| 37 | 82.5 | 5.2 | 624 | 12 | US-10-282-122A-51658 | Sequence 51658, A |
| 38 | 82.5 | 5.2 | 854 | 14 | US-10-244-488-13     | Sequence 13, Appl |
| 39 | 82.5 | 5.2 | 996 | 15 | US-10-369-493-18505  | Sequence 18505, A |
| 40 | 81.5 | 5.1 | 275 | 12 | US-10-425-114-69783  | Sequence 69783, A |
| 41 | 81   | 5.1 | 329 | 15 | US-10-369-493-233    | Sequence 233, App |
| 42 | 81   | 5.1 | 645 | 12 | US-10-282-122A-53367 | Sequence 53367, A |
| 43 | 81   | 5.1 | 905 | 13 | US-10-114-893-127    | Sequence 127, App |
| 44 | 81   | 5.1 | 984 | 15 | US-10-369-493-20751  | Sequence 20751, A |
| 45 | 80   | 5.0 | 275 | 12 | US-10-282-122A-52681 | Sequence 52681, A |

ALIGNMENTS

RESULT 1  
US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV.  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4

Query Match 100.0%; Score 1597; DB 9; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.3e-171;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLDFPEWERCCKPLTRLHVTYEGTIENCGMLQV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTRQSLDFPEWERCCKPLTRLHVTYEGTIENCGMLQV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISLFTVLDHNECLIITGTQEYSEYTGYA 120

Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGYA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 916  
QY 241 FLTERKLTVDVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTVDVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976

RESULT 2

US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

Query Match 93.7%; Score 1497; DB 9; Length 968;  
Best Local Similarity 93.7%; Pred. No. 2.9e-160;  
Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTQSLDEDFPEWERCCKPLTRLHVITYEGTIEENGQGLQV 60  
Db 669 KTLFCYFRRVTEKKPTGLVTFTQSLDEDFPEWERCCKPLTRLHVITYEGTIEENGQGLQV 728  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGIA 120  
Db 729 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGIA 788  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 789 ETYRWSRSHEDGSEKDDWQRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 848  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 849 VPSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHT 908  
QY 241 FLTERKLTVDVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 909 FLTERKLDVGKVKLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 968

RESULT 3

US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

Query Match 93.5%; Score 1493; DB 9; Length 977;  
Best Local Similarity 94.0%; Pred. No. 8.3e-160;  
Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTQSLDEDFPEWERCCKPLTRLHVITYEGTIEENGQGLQV 60  
Db 678 KTLFCYFRRVTEKKPTGLVTFTQSLDEDFPEWERCCKPLTRLHVITYEGTIEENGQGLQV 737  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGIA 120  
Db 738 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNECLIITGTEQYSEYTGIA 797  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 798 ETYRWSRSHEDGSRDDWQRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLRPG 857  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 858 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHT 917  
QY 241 FLTERKLTVDVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 299  
Db 918 FLTERKLTVDVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESCTQITNQPGQRTGT 976

RESULT 4

US-09-973-451-8  
; Sequence 8, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster

US-09-973-451-8  
; FEATURE:

|                       |                  |                  |            |             |
|-----------------------|------------------|------------------|------------|-------------|
| Query Match           | 38.8%;           | Score 619.5;     | DB 9;      | Length 768; |
| Best Local Similarity | 46.4%;           | Pred. No. 1e-60; |            |             |
| Matches 140;          | Conservative 41; | Mismatches 90;   | Indels 31; | Gaps 8;     |

  

|    |     |   |                             |              |     |
|----|-----|---|-----------------------------|--------------|-----|
| QY | 1   | KTLCFCYFRRV--TEKK----   | PTGLVTFTRQS--LEDFPWECEKPL-- | TRLHVTVYEGTI | 50  |
|    |     | :   | :     :                     |              |     |
| Db | 297 | KCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPEHLIDWSQSAAPGLGDVPLHVDAGETI |                             |              | 356 |
|    |     | :   | :     :                     |              |     |
| QY | 51  | EENGQQLQVDFANRFVGGVTSAGLVQEEIRFLINPELLISRLFTFVLHDHNECLII      |                             | TGT          | 110 |
|    |     | :   | :     :                     | :            |     |
| Db | 357 | EDEGIGLLQVDFANKYLGCGVLGHGCVQEEIRFVICPELLVKGKLFTECLRPFEALVMLGA |                             | 416          |     |
|    |     | :   | :     :                     | :            |     |
| QY | 111 | EQSEYTGAYETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELN    |                             | 170          |     |
|    |     | :   | :     :                     | :            |     |
| Db | 417 | ERYSNYTGAGSPFWSGNFEDSTPRDSSGRRQTAVAIIDALHFAQSHHQYREDLMERELN   |                             | 476          |     |
|    |     | :   | :     :                     | :            |     |
| QY | 171 | KAYCGFLR-----PGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYF   |                             | 224          |     |
|    |     |   |                             | :     :      |     |
| Db | 477 | KAYIGFVHMVWVTPPPG-----VATGNWCGGAFGGDSYLLKALLQMLVCAQLGRPLAYY   |                             | 528          |     |
|    |     |   |                             | :     :      |     |
| QY | 225 | TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLRLRYNEECRNCTPGP-----DIKLYP    |                             | 278          |     |
|    |     | :     :   | :     :                     |              |     |
| Db | 529 | TFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRSYSRLIKESSKEPRENKASKKKLYD   |                             | 587          |     |
|    |     | :     :   | :     :                     |              |     |
| QY | 279 | FI 280  |                             |              |     |
|    |     |   |                             |              |     |
| Db | 588 | FI 589  |                             |              |     |
|    |     |   |                             |              |     |

## RESULT 5

```

US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-262-E12_FLI.pep
US-10-425-114-60000

```

|                           |       |                  |            |             |
|---------------------------|-------|------------------|------------|-------------|
| Query Match               | 30.6% | Score 489;       | DB 12;     | Length 546; |
| Best Local Similarity     | 38.4% | Pred. No. 4e-46; |            |             |
| Matches 113; Conservative | 46;   | Mismatches 101;  | Indels 34; | Gaps 5;     |

```

QY      1  KTLFCYFRRRVTEKKPTGLVTFTRQSL-----EDFPWEWERCCKPLTRLHLVITYEGT  49
Db      226 RCLVHYFERVTDSMTPGLVSPERKVLPRRALSDGVYPDIHAWVASSAPLQCFVTFSSGF  285

QY      50  IEENGQGMQLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLHDNECLIITG  109
Db      286 IEDEEQEALQVDFANKYLGGAALSWGCVQEEIRFMINPELIVGMLFLSCMEDNEAIEIFG  345

QY     110  TEQYSEYTGYAETVRSRSHEDGSEDDCERRCTEIVAIDALHFRRYLDQFVPEKMRREL  169
Db     346 AERFSQYMGYGSSFRFVGVDLDTKPFDSMGRRRTRIVAIDALDCPARL-HYESGCLLRVY  404

```

```

QY 170 NKAYCGFLR-----PGVSSSENLSAVATGNWGCAGFGGDARLKALI 209
Db 405 NKAFCGFFDQSKHLYAKLFDQLHNKDDFSSINSSEYVGSTGNWGCAGFGGNPEIKSMI 464

QY 210 QILAAAAAERDVV-YFTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLRLRYNE 262
Db 465 QWIAASQALRPFNYYTFEDVSLQR-LEEVIQWIRLHGWTVGELWHMLMEYSSQ 517

```

## RESULT 6

US-09-973-451-10  
; Sequence 10, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA2 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10

|                       |                  |                    |           |             |
|-----------------------|------------------|--------------------|-----------|-------------|
| Query Match           | 23.3%            | Score 371.5;       | DB 9;     | Length 726; |
| Best Local Similarity | 35.3%;           | Pred. No. 1.4e-32; |           |             |
| Matches 94;           | Conservative 49; | Mismatches 104;    | Indels 19 |             |

```

QY 1 KTLFCYFRRVTEKKPTGLVTF--TRQSEDF-PEWERCEKPLTRLHVT--YEGTIEENGQ 55
Db   ||| ||| ::| ||| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::|
QY 456 KFLFTYFDRKNSMDPPDGAVSFRLTKMDKDTFNEEWK--DKKLRSPLPEVEFFDEMLIEDTA 513
Db   ||| ||| ::| ||| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::|
QY 56 GMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELLIISRLFTVLDHNECLIITGTQYSE 115
Db   ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::|
QY 514 LCTQVDFANEHLGGGVLNHGSVQEEIRFLMCPENVMGMLCEKMKQLEAISIVGAYVFSS 573
Db   ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::|
QY 116 YTGAAETVRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFRR-----YLDQFVPEK 164
Db   ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::|
QY 574 YTGYGHTLKWAELOPNHSRQNTNEFRDRFGRLVRVETIADAILFKGSKLDCQTEQLNKAN 633
Db   ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::| ||| ||| ::|
QY 165 MRREINKAYCGFLRPGVSSSENLSAVATGNWGGCAFGGDARLKALIQILAAAAAERDVVVF 224
Db   : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGAFNGDKPLKFLIIQIVIAAGVADRPLHFC 692
Db   : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250
Db   : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 693 SFGEPELAAKCKKIIERMKOKDVTLG 718
Db   : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```

## RESULT 7

US-10-424-599-184988  
; Sequence 184988, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(200)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138059C.1.pep
US-10-424-599-184988

Query Match          12.0%; Score 191.5; DB 12; Length 200;
Best Local Similarity 36.0%; Pred. No. 5.6e-13;
Matches 40; Conservative 19; Mismatches 41; Indels 11; Gaps 2;

QY      6 YFRRVTEKPTGLVTFTRQSL-----DFPE---WERCEKPLTRLHVTYEGTIEENG 54
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      90 YQRISSSEMPKGVISFERKVLFPFKNDSIHISYPDANFWSTSAIPLCRFEVHSSGLIEDQS 149
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      55 QGMLQVDFANFVGGVTSAGLVQEEIRFLINPELLIISRLFTTEVLHDNECL 105
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      150 SGAAEVDLANKYLGCGALGRGCVQEEIRFMVSPLEAGMLFLPAMADNEAI 200
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 8
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445

Query Match          11.8%; Score 188; DB 12; Length 180;
Best Local Similarity 30.1%; Pred. No. 1.2e-12;
Matches 49; Conservative 23; Mismatches 41; Indels 50; Gaps 4;

QY      168 ELNKAYCGFL-----RP----- 179
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      7 EINKAFCGFLYQCKYQPYQKILQENGCTSAIFYAATSTSMETDEGEISNHKITSQNDYH 66
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      180 GVSSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERD-VVYFTFGDSLMRDIYSM 238
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      67 GMDQGNNIGVATGNWCGAFGGDPEVKTIIQWLAAASQALRPFIAYTTFG-LEALQSLDEV 125
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      239 HIFLTERKLTVDVYKLLRLRYNEECRNCSTPGPDIKLYPFIY 281
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      126 AHWILSQRTVGDGLWNMLIEYSINRSKGETNVGFLQWLLPSIY 168
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 9
US-09-973-451-11
; Sequence 11, Application US/09973451
; Patent No. US20020132328A1
```

```
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 11
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; US-09-973-451-11

Query Match          10.5%; Score 167; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      94 LFTTEVLHDHNECLITGTQYSEYTGAEYR 124
      |||||
Db      1 LFTTEVLHDHNECLITGTQYSEYTGAEYR 31
      |||||

RESULT 10
US-09-973-451-12
; Sequence 12, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 12
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; US-09-973-451-12

Query Match          9.3%; Score 149; DB 9; Length 29;
Best Local Similarity 96.6%; Pred. No. 2.1e-09;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      172 AYCGLRPGVSSSENLSAVATGNWCGAFG 200
      |||||
Db      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
      |||||

RESULT 11
US-10-424-599-233915
; Sequence 233915, Application US/10424599
```



; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 233915  
; LENGTH: 300  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_5324C.1.pep  
US-10-424-599-233915

Query Match 6.2%; Score 99.5; DB 12; Length 300;  
Best Local Similarity 21.4%; Pred. No. 0.029;  
Matches 36; Conservative 35; Mismatches 56; Indels 41; Gaps 5;  
QY 9 RVTEKKPTGLVTFTRQSLED-----FPEWER--CEKPLTRLHVTYEGTIENGQGM 57  
Db 133 RLIDSQPGGIVFLTQELIAALLSCSLFCLFPVSDRPVHLPMINFDVLF-----GS 183  
QY 58 LQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIITGTEQYSEYT 117  
Db 184 LYDDYSQK-----QENKIWCIVHYFQRISSSEMPKGIAYNERIEIVGVVERFSGYT 232  
QY 118 GYAETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRLDQFVPEKM 165  
Db 233 DHASSFRFSRAKAEGRDPVGRKTSDL-----EKYFPPKM 270

RESULT 12  
US-10-282-122A-51256  
; Sequence 51256, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 51256  
; LENGTH: 395  
; TYPE: PRT  
; ORGANISM: Bordetella pertussis  
US-10-282-122A-51256  
Query Match 6.1%; Score 97.5; DB 12; Length 395;  
Best Local Similarity 24.3%; Pred. No. 0.074;  
Matches 54; Conservative 25; Mismatches 76; Indels 67; Gaps 9;  
QY 73 SAGLVQEEIRFLINP-----ELIISRLFTFVLDHNECLIITGTEQYSEYTYGAETYRW- 125  
Db 75 NAGICKETPAFNVNRLCGSLQAIVSAQAOTIMLGDTEIAGAGAESMSRAPYIAPAQRWG 134  
QY 126 -----SRSH-----EDGERDDCERRCTEIVAID-----ALHF 153  
Db 135 ARMGDSAMVDMTGTALSDPFGRMHMGVTAENVAAARYSISRADQDALAVESHRAAAIDA 194  
QY 154 RRYLDQFVP-----EKMRELKAYCGFLRPGVSSENLSAVATGNWCGGA 198  
Db 195 GRFRDQIVPVLTKRKGETVFTDTEHVRRDITIEGMAALRPVFOKEN-GTVTAGN----A 249  
QY 199 FGGDARLKALIIILAAAAAARD-----VVYFTFG-DSELM 232  
Db 250 SGLNDGAGAVVLMSSAAAAAQRGIAPMARLVAYAHAGVDPPEIM 291

RESULT 13  
US-10-369-493-19591  
; Sequence 19591, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 19591  
; LENGTH: 730  
; TYPE: PRT  
; ORGANISM: Nitrosomonas europaea  
US-10-369-493-19591

Query Match 5.9%; Score 94.5; DB 15; Length 730;  
Best Local Similarity 22.0%; Pred. No. 0.4;  
Matches 68; Conservative 46; Mismatches 96; Indels 99; Gaps 17;  
QY 8 RRVTEKK-PTGLVTFTRQSLE-----DPPEWERCEKPLTRLHVTYEGTIEENG 54  
Db 219 RRITENRVDPDLANHQVYALDMGALLAGTKYRGFEQ-----RLKVVLK-QLTDNP 268  
QY 55 QGMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIITGTEQYS 114  
Db 269 KAILFIDEIHTLIGAGAASGGTL--DASNLKLPALSSGRL-----KCI---GATTYN 315  
QY 115 EYTG-YAETYRWSRSHEDGSRDDCER-----RCTEI--VAIDALH 152

Db 316 EYRGVFEKDHLSRRPFQKIDVSEPDIGETVEILRLGLKSRYEKHNVKYTEVALTAAELS 375  
Qy 153 FRRYLDQFVPEKM-----RRELNKAYCGFLR---PGVSSSEN 185  
Db 376 ARFINDRHLPDKAIDVIDEAGAAQRILPKSRQRKIIGRQEIIEQVIAGIARIPPQNVSSDD 435  
Qy 186 LSAVATGNWCGA--FGGDARLKALIQILAAAAAE-----RDVVYFTF-----GDSELM 232  
Db 436 RNKLTLDRLDKAIVFGQDAADALTSAIKMARSLGLNTCKPIGSFLPSGPTGVGKTEVA 495  
Qy 233 RDI-YSMHI 240  
Db 496 RQLAYTLGI 504

RESULT 14  
US-10-314-657-46  
; Sequence 46, Application US/10314657  
; Publication No. US20030175888A1  
; GENERAL INFORMATION:  
; APPLICANT: SHEN, Ben  
; APPLICANT: CHENG, Yi-Qiang  
; APPLICANT: TANG, Gong-Li  
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide  
; TITLE OF INVENTION: Synthases and Methods of Use  
; FILE REFERENCE: 054030-0021  
; CURRENT APPLICATION NUMBER: US/10/314,657  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: PCT/US02/08937  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US 60/278,935  
; PRIOR FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 214  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 46  
; LENGTH: 7349  
; TYPE: PRT  
; ORGANISM: Streptomyces atroolivaceus  
US-10-314-657-46

Query Match 5.8%; Score 92.5; DB 14; Length 7349;  
Best Local Similarity 22.5%; Pred. No. 21;  
Matches 62; Conservative 31; Mismatches 89; Indels 93; Gaps 12;  
Qy 40 TRLHVTYE--GTIEENGQMLQVDFANRFVGGVTSAGLVQBEI-----RFLI 85  
Db 641 TRLTVRHETLAVPSGNGAGVREDGVYLVGTGGGSLAALLVDRLVTRGVPVRLVLTGRSAP 700  
Qy 86 NPELI-----ISRLFTEVLDHNECL--IITGTEQYSEYTGyaETVWRSRSHED 131  
Db 701 GPFLTQRIEGWRRRGAEVTHVRGDVAHTDDVLAAVTCARETYGRIDG---VFHCAGSVDD 757  
Qy 132 GS-ERDDCERRCT-----EIVADALHF-----RR 155  
Db 758 GMFFRKDPERSAAVLAAKVAGTRNLDEATADDGLAFALFSSVSASVANPGQADYAYGNA 817  
Qy 156 YLDQFVPEKMRRELNKAYCGFLRPGVS-----SENLSAVATGNWCGGAFG 200  
Db 818 FMEHFAEQRAARA-----DRPGVSVAGWPLWADGGMRVSEDLRRSADTSGLHALP 869  
Qy 201 GDARLKALIQILAAAAAERDVVYFTFGDSELMRDI 235  
Db 870 ADAGLDALFGLL-SGAAPRAVV--TYGDQERIAEL 901

RESULT 15  
US-10-247-671-134  
; Sequence 134, Application US/10247671  
; Publication No. US20030194721A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikita, Thomas  
; APPLICANT: Shiffman, Dov  
; APPLICANT: Porter, Gordon, J.

; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
; FILE REFERENCE: PA-0050 US  
; CURRENT APPLICATION NUMBER: US/10/247,671  
; CURRENT FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: 60/323,784  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 186  
; SOFTWARE: PERL Program  
; SEQ ID NO 134  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030194721A1 2616727CD1  
US-10-247-671-134  
Query Match 5.7%; Score 91.5; DB 14; Length 261;  
Best Local Similarity 22.6%; Pred. No. 0.19;  
Matches 53; Conservative 32; Mismatches 80; Indels 69; Gaps 9;  
Qy 12 EKKPTGLVTFTRQS---LEDFFPEWERC-EKPLTRLHVTYEGTIEEN-GQGMLQVDFANRF 66  
Db 48 EKKPYCNAHYPKQSFTMVADTTPENLRKQQSELSQVRYKEEFKXKKGK-----97  
Qy 67 VGGGVTSAGLVQBEIRFLINPELIISRLFTEVLDHNECLIIITGTEQYSEYTGyaETVWRWS 126  
Db 98 -----FSVVADTPELQRIKKTQDQISNIKYHEEFKS 129  
Qy 127 RSHEDGSEDDCERRCTEIVADALHFRRYLDQFVP-----EKMRELNKAYCG 175  
Db 130 RMGPSGGEGMEPERRDSQ---DGSSYRPLEQQQPHHIPTSAVYQQPQQPVAQSYGG 185  
Qy 176 FLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDS 229  
Db 186 YKEP-AAPVSIQRSAPG-----GGGKRYRAVYDY---SAADEDEVSFQDGD 228  
Search completed: May 26, 2004, 19:18:56  
Job time : 26.0302 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 8.25272 Seconds  
(without alignments)  
3496.721 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977  
Perfect score: 1597  
Sequence: 1 KTLFCYFRRVTEKKPTGLVT.....YHAVESCAETADHSGQRTGT 300

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID    | Description        |
|------------|-------|-------|--------|----------|--------------------|
| 1          | 406   | 25.4  | 997    | 2 B84726 | probable poly(ADP- |
| 2          | 371.5 | 23.3  | 726    | 2 T21138 | hypothetical prote |
| 3          | 254.5 | 15.9  | 364    | 2 A84726 | probable poly(ADP- |
| 4          | 91.5  | 5.7   | 261    | 2 S68234 | lasp-1 protein - h |
| 5          | 91.5  | 5.7   | 491    | 2 S05408 | keratin, type II,  |
| 6          | 90    | 5.6   | 1194   | 2 S70415 | DNA-directed RNA p |
| 7          | 88.5  | 5.5   | 2135   | 2 T14602 | variant-specific B |
| 8          | 88    | 5.5   | 448    | 2 AB2740 | acetyl-CoA carboxy |
| 9          | 88    | 5.5   | 448    | 2 H97520 | biotin carboxylase |
| 10         | 87.5  | 5.5   | 391    | 2 AB2322 | hypothetical prote |
| 11         | 86.5  | 5.4   | 153    | 2 G90831 | probable endopepti |
| 12         | 86    | 5.4   | 153    | 1 APBPML | endopeptidase (EC  |
| 13         | 86    | 5.4   | 153    | 1 APBP21 | endopeptidase (EC  |
| 14         | 85    | 5.3   | 523    | 2 S03572 | DNA-directed RNA p |
| 15         | 85    | 5.3   | 523    | 2 B84416 | DNA-directed RNA p |
| 16         | 84.5  | 5.3   | 317    | 2 F82672 | ATP sulfurylase, s |
| 17         | 84.5  | 5.3   | 397    | 2 T35713 | probable oxidoredu |
| 18         | 83.5  | 5.2   | 153    | 2 B64788 | endopeptidase ybct |
| 19         | 83.5  | 5.2   | 3036   | 2 T18995 | hypothetical prote |
| 20         | 83    | 5.2   | 729    | 2 A69202 | conserved hypothet |
| 21         | 82.5  | 5.2   | 624    | 2 D97307 | molecular chaperon |
| 22         | 82.5  | 5.2   | 681    | 2 S27868 | glvr-1 protein - m |
| 23         | 82.5  | 5.2   | 996    | 2 D86872 | beta-galactosidase |
| 24         | 82    | 5.1   | 256    | 2 S16255 | hypothetical prote |
| 25         | 82    | 5.1   | 282    | 2 A97214 | uncharacterized co |
| 26         | 81.5  | 5.1   | 600    | 2 A45112 | major paraflagella |
| 27         | 81    | 5.1   | 602    | 2 AH2583 | ATP-dependent DNA  |
| 28         | 81    | 5.1   | 602    | 2 H97365 | DNA helicase XF138 |
| 29         | 80.5  | 5.0   | 477    | 2 C69318 | reductase, iron-su |

|    |      |     |     |          |                    |
|----|------|-----|-----|----------|--------------------|
| 30 | 80.5 | 5.0 | 576 | 2 AC2195 | hypothetical prote |
| 31 | 79.5 | 5.0 | 618 | 2 G69114 | indolepyruvate oxi |
| 32 | 79.5 | 5.0 | 723 | 2 T47613 | ABC transporter-li |
| 33 | 79   | 4.9 | 363 | 2 T01354 | herbicide safener  |
| 34 | 79   | 4.9 | 502 | 2 T26256 | hypothetical prote |
| 35 | 78.5 | 4.9 | 418 | 2 G95231 | competence-induced |
| 36 | 78.5 | 4.9 | 459 | 2 A71840 | phosphohexosemuta  |
| 37 | 78.5 | 4.9 | 761 | 1 S61642 | translation elonga |
| 38 | 78.5 | 4.9 | 865 | 2 H71447 | trehalose-6-phosph |
| 39 | 78.5 | 4.9 | 990 | 1 SNFFIN | insulysin (EC 3.4. |
| 40 | 78   | 4.9 | 412 | 2 T46104 | hypothetical prote |
| 41 | 78   | 4.9 | 469 | 2 T41723 | serine/threonine-s |
| 42 | 78   | 4.9 | 665 | 1 T44793 | beta-galactosidase |
| 43 | 78   | 4.9 | 665 | 2 S70706 | probable protein k |
| 44 | 78   | 4.9 | 670 | 1 H70421 | hypothetical prote |
| 45 | 78   | 4.9 | 697 | 2 T21567 | hypothetical prote |

ALIGNMENTS

RESULT 1

B84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84726  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-997 <STO>  
A;Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g31870  
A;Map position: 2

|                       |        |  |       |                |
|-----------------------|--------|--|-------|----------------|
| Query Match           | 25.4%; | Score 406;   | DB 2; | Length 997;    |
| Best Local Similarity | 34.5%; | Pred. No. 2.9e-29;   |       |                |
| Matches               | 105;   | Conservative   | 33;   | Mismatches 72; |
| Indels                | 94;    | Gaps   | 9;    |                |
| QY                    | 6      | YFRRVTEKKPTGLVTFTRQ--SLEDFPEWERCERKPLTRLHVITYEGTIEENGQGLQVDEFA | 63    |                |
| Db                    | 196    | YFERFCSCVPIGIVSFERKITAAPDADFWSKSDVSL-----YQ-----PDNALEVDFA     | 243   |                |
| QY                    | 64     | NRFVGGVTSAGLVQHEIRFLINPELIISRLFTFVLDHNECLIIITGTEQYSEYTGVAETY   | 123   |                |
| Db                    | 244    | NKYLGGGSLSRGCVQBEIRFMINPELIAGMLFLPRMDDNEAIEIVGAERFSCYTGVASSF   | 303   |                |
| QY                    | 124    | RWSRSHEDGSSERDDCERRCTEIVAIDAL-----HFRRYLDQFVPEKMRRELNKAYCGFL   | 177   |                |
| Db                    | 304    | RFAGEYIDKKAMPFKRRTRIVAIDALCTPKMRHFKDIC-----LLREINKALCGFL       | 356   |                |
| QY                    | 178    | R-----   | 178   |                |
| Db                    | 357    | NCSKAWEHQNI FMDEGDNEIQLVRNGRDSGLLRTETTASHTPLNDVEMREKPPANNLIR   | 416   |                |
| QY                    | 179    | -----PGVSSSEN--SAVATENWCGGAFGGDARLKALIQILAAAAAERDVV-YFTFGDSEL  | 231   |                |
| Db                    | 417    | DFYVEGVDNEDHEDDGVATGNWCGVFGGDPDLKATIQWLAASQTRRRPFISYTFG-VEA    | 475   |                |
| QY                    | 232    | MRDI 235   |       |                |
| Db                    | 476    | LRNL 479   |       |                |

RESULT 2

T21138

hypothetical protein F20C5.1 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C/Accession: T21138  
R/Matthews, P.  
submitted to the EMBL Data Library, December 1995  
A/Reference number: Z19381  
A/Accession: T21138  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-726 <WIL>  
A/Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1  
A/Experimental source: clone F20C5  
C/Genetics:  
A/Gene: CESP:F20C5.1  
A/Map position: 4  
A/Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1  
C/Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1

Query Match 23.3%; Score 371.5; DB 2; Length 726;  
Best Local Similarity 35.3%; Pred. No. 3e-26;  
Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRRVTEKKPTGLVTF--TRQSLDF-PEWERCEKPLTRLHVT--YEGTIEENGQ 55  
Db 456 KPLFTYFDKMSMDPPDGAVSPLTKMDKDTFNEEWK--DKLRSLPEVEFFDEMLIEDTA 513  
QY 56 GMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNECLIITGTREQYSE 115  
Db 514 LCTQVDFANEHLGGVNLHGVSQEEIRFLMCPENMVGLCEKMKQLEAISIVGAYVFSS 573  
QY 116 YTGVAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFR-----YLDQFVPEK 164  
Db 574 YTGYGHTLKWAELOPNHRSQNTNEFRDRFGLRVETIATDAILFKGSKLDCQTEQLNKAN 633  
QY 165 MRRELNKAYCGFLRPGVSSENLSAVATGNWCGAGFGGDARLKALIQILAAARRDVVYF 224  
Db 634 IIREMKKASIGFMSQGPKFNTI-PIVTGWGCGAFNGDKPLKFIIQIVIAAGVADRPLHFC 692  
QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250  
Db 693 SFGEPELAAKCKKIIERMKQKDVTLG 718

RESULT 3  
A84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C/Accession: A84726  
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: A84726  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-364 <STO>  
A/Cross-references: GB:AE002093; NID:G4887750; PIDN:AAD32286.1; GSPDB:GN00139  
C/Genetics:  
A/Gene: At2g31860  
A/Map position: 2

Query Match 15.9%; Score 254.5; DB 2; Length 364;  
Best Local Similarity 28.3%; Pred. No. 9.5e-16;  
Matches 76; Conservative 36; Mismatches 78; Indels 79; Gaps 7;

QY 1 KTLFCYFRRVTEKKPTGLVTFTRQ--SLE-----DFPEWERCEKPLTRLHVTYEGTIEE 52  
Db 112 KCIHYFORLSSSISPGFVSFERKILSLEQDSSTLDEGFWGKSTVNLCPVEVTSGLIED 171

QY 53 NGQGMLOVDFAANRFVGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNECLIITGTTEQ 112  
Db 172 QSVEALEVDFANKNLGGGALRKGCVQEEIRFMINPELVGMFLPTMEVTEAIEVVGAE 231  
QY 113 YSEYTGVAETYRWSRSHEDGSDRDCERRC-TEIVAIDALHFERYLDQFVPEKMRRELNK 171  
Db 232 FSLYTG-----CFRKAKTIVAIDALR----- 253  
QY 172 AYCGLRPGVSSENLSAVATGNWCGAGFGGDARLKALIQILAAARRDVVYF----- 225  
Db 254 -----HPGVS-----QYKLESLLSVLILSSSGRPRLYMGVSLSLQ 288  
QY 226 -FGDSELMRDIYSMHIFLTERKLTVGDVY 253  
Db 289 GIGDVVLMVEILSSSLFFENGLRFRHSNLY 317

RESULT 4  
S68234  
Lasp-1 protein - human  
N/Alternate names: LIM and SH3 protein  
C/Species: Homo sapiens (man)  
C/Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 16-Jul-1999  
C/Accession: S68234; S72556  
R/Tomasetto, C.; Moog-Lutz, C.; Regnier, C.H.; Schreiber, V.; Basset, P.; Rio, M.C.  
FEBS Lett. 373, 245-249, 1995  
A/Title: Lasp-1 (MLN 50) defines a new LIM protein subfamily characterized by the assoc.  
A/Reference number: S68234; MUID:96033982; PMID:7589475  
A/Accession: S68234  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-261 <TOM>  
A/Cross-references: EMBL:X82456  
R/Tomasetto, C.  
submitted to the EMBL Data Library, November 1994  
A/Reference number: S72556  
A/Accession: S72556  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-84,'C',86-90,'G',92-219,'D',221-261 <TO2>  
A/Cross-references: EMBL:X82456; NID:951273; PID:951274  
C/Superfamily: LIM metal-binding repeat homology; SH3 homology  
F;5-56/Domain: LIM metal-binding repeat homology <LIM>  
F;209-258/Domain: SH3 homology <SH3>

Query Match 5.7%; Score 91.5; DB 2; Length 261;  
Best Local Similarity 22.6%; Pred. No. 0.92;  
Matches 53; Conservative 32; Mismatches 80; Indels 69; Gaps 9;

QY 12 EKKTGLVTFTRQS---LEDFPWERC-EKPLTRLHVTYEGTIEEN-GQGMLQVDFANRF 66  
Db 48 EKKPYNCAHYPKQSFTMWADTPENLRKQSQBELQSVRYKEEFKNKGK----- 97  
QY 67 VGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNECLIITGTQYSEYTGVAETYRWS 126  
Db 98 -----FSVVADTPELQRIKKTQDQISNIKYHEEFKS 129  
QY 127 RSHEDGSDRDCERRCTEIVAIDALHFRRYLDQFVP-----EKMRRELNKAYCG 175  
Db 130 RMGPSGGEGMEFERDSQ----DGSSYRPLEQQQPHIPTSAVYQQPQQPVAQSYGG 185  
QY 176 FLRPGVSSENLSAVATGNWCGAGFGGDARLKALIQILAAARRDVVYFTFGDS 229  
Db 186 YKEP-AAPVSIQRSAPG-----GGGKRYRAYDY---SAADEDEVSFQDGD 228

RESULT 5  
S05408  
keratin, type II, component 7c, cytoskeletal - sheep (tentative sequence)  
C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C/Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 31-Mar-2000  
C/Accession: S05408  
R/Sparrow, L.G.; Robinson, C.P.; McMahon, D.T.W.; Rubira, M.R.





Db 373 YKIPPYDSMIGKLVHGRDRDECIRR-----LRRALDEFVVDGIKTTL 416

RESULT 9

H97520

biotin carboxylase (a chain of acetyl-CoA carboxylase (acc) [imported] - Agrobacterium tumefaciens

C;Species: Agrobacterium tumefaciens

C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002

C;Accession: H97520

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens

A;Reference number: A97359; PMID:21608551; PMID:11743194

A;Accession: H97520

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-448 <KUR>

A;Cross-references: GB:AE007869; PIDN:AAK87121.1; PID:g15156385; GSPDB:GN00169

C;Genetics:

A;Gene: AGR\_C\_2451

A;Map position: circular chromosome

C;Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match 5.5%; Score 88; DB 2; Length 448;

Best Local Similarity 27.4%; Pred. No. 4;

Matches 48; Conservative 16; Mismatches 59; Indels 52; Gaps 9;

QY 33 ERCEKPLTRLHVITYEGTIE---ENGQ-----GMLQVDFANRFVGGVTSAGLVQEEIR 82

Db 256 EICADAMRKLKYGAGTIFLYENGGEFYFIEMNTRLQVEHP---VTEAITGMDLVQEQIR 312

QY 83 FLINPELIISRLFTFVLNHN-ECLII-----ITGTEQY-----SEYTG 118

Db 313 VASGQGLSVTQADIEFHGHAIECRINAEDPRTFVPSGTLTYFHTPGGLGVRVDSGAYQG 372

QY 119 Y-AETRWRSRSH---DGSERDDCERRCTEIVAIDALHFRYLDQFVPEKMRREL 169

Db 373 YKIPPYDSMIGKLVHGRDRDECIRR-----LRRALDEFVVDGIKTTL 416

RESULT 10

AB2322

hypothetical protein alr4129 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AB2322

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A;Reference number: AB1807; PMID:21595285; PMID:11759840

A;Accession: AB2322

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-391 <KUR>

A;Cross-references: GB:BA000019; PIDN:BAW75828.1; PID:g17133264; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr4129

C;Superfamily: hypothetical protein AF0390

Query Match 5.5%; Score 87.5; DB 2; Length 391;

Best Local Similarity 20.5%; Pred. No. 3.7;

Matches 50; Conservative 33; Mismatches 96; Indels 65; Gaps 10;

QY 36 EKPLTRLHVITYEGTIE---ENGQMLQVDFANRFVGGV-----TSAGLVQEEIRFLINPE 88

Db 140 EYPQIHILHAFSPQEQVFARVDGLEIVDVVISALQNAGVNSLPGTAAEVLDDVEIRILCPE 199

QY 89 LIISRLFTFVLN-----HNECLIIITG-TEQYSEYTGVAETRWRSRSHEDGSRDDCER 140

Db 200 KINTATWLEIIGTAKVKVGLHTTSTILSCHIEPTQKIGHLEKL---RSLQQTAINHKYPA 256

QY 141 RCTEIVAIDALHFRYLDQFVPEKMRRELN-----KAYCGFLRPG----- 180

Db 257 RITEFIVLP-----FVQGEAPKSLRRRVGRDQIPILADALLGAVARIYLGNTWIPNHQPS 310

QY 181 -VSSENLSAVATGNWGC-----GAFGGD-----ARLKALIQILAAAAAER 219

Db 311 WVKLGLAGATEALTWGCNDIGGTLMEEHITTMAGAVGGTCMEVATLQNAIASIGRPYQQR 370

QY 220 DVVY 223

Db 371 DTLY 374

RESULT 11

G90831

probable endopeptidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05)

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C;Accession: G90831

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic

A;Reference number: A99629; PMID:21156231; PMID:11258796

A;Accession: G90831

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-153 <HAY>

A;Cross-references: GB:BA000007; PIDN:BAW35046.1; PID:g13361087; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECs1623

C;Superfamily: phage PA2 endopeptidase

Query Match 5.4%; Score 86.5; DB 2; Length 153;

Best Local Similarity 23.7%; Pred. No. 1.3;

Matches 46; Conservative 24; Mismatches 59; Indels 65; Gaps 9;

QY 90 IISRLFTFVLNHECLIIITGTEQYSEYTGVAETRWRSRSHEDGSRD-----DCER 140

Db 7 IISALLI-----CIIVCLSWAVNHYRDNAITYKAQR---DKNARELKLANVAITDMQM 56

QY 141 RCTEIVAIDALHFRYLDQFVPEKM-----RRELN-KAYCGFLRPGVSSSENLSAVAT 191

Db 57 RQRDVAALDAKYTKELADAKAENDALHDDVAAGRRLHIKAVCQSVREATTASGVDNAAS 116

QY 192 GNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTIVGD 251

Db 117 -----PRL-----ADTAERD--YFT-----LRERLITMQK 139

QY 252 VYKLLRLRYNNEECR 265

Db 140 QLEGTKYINEQCR 153

RESULT 12

APBPML

endopeptidase (EC 3.4.-.-) - phage lambda

C;Species: phage lambda

C;Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 18-Jun-1999

C;Accession: A94614; C92891; JN0749; A00906

R;Daniels, D.

submitted to the Nucleic Acid Sequence Database, September 1982

A;Reference number: A94614

A;Accession: A94614

A;Molecule type: DNA

A;Residues: 1-153 <DAN>

R;Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B. J. Mol. Biol. 162, 729-773, 1982

A;Title: Nucleotide sequence of bacteriophage lambda DNA.

A;Reference number: A92891; PMID:83189071; PMID:6221115

A;Accession: C92891  
A;Molecule type: DNA  
A;Residues: 1-153 <SAN>  
A;Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:9215104;  
R;Hanych, B.; Kedzierska, S.; Walderich, B.; Uznanski, B.; Taylor, A.  
Gene 129, 1-8, 1993  
A;Title: Expression of the Rz gene and the overlapping Rz1 reading frame present at the  
A;Reference number: JN0749; MUID:93328108; PMID:8335247  
A;Accession: JN0749  
A;Molecule type: DNA  
A;Residues: 1-153 <HAN>  
C;Comment: Gene Rz is necessary for host cell lysis. It is believed to code for an endop-  
he murein component of the bacterial cell wall.  
C;Genetics:  
A;Gene: Rz  
A;Map position: 94.77-95.72  
C;Superfamily: phage PA2 endopeptidase  
C;Keywords: host cell lysis; hydrolase

|    | Query Match           | 5.4%;                       | Score 86;                | DB 1;           | Length 153;            |
|----|-----------------------|-----------------------------|--------------------------|-----------------|------------------------|
|    | Best Local Similarity | 23.3%;                      | Pred. No. 1.5;           |                 |                        |
|    | Matches 42;           | Conservative                | 24;                      | Mismatches 56;  | Indels 58; Gaps 8;     |
| QY | 104                   | CLIIITGEQYSEYTGVAETRWRS     | SHEDG                    | SERD-----       | DCERRCTEIVAIDALHFR 154 |
|    |                       | ::: :         :       :     |                          |                 |                        |
| Db | 14                    | CIIVCLSWAVNHYRDNAITYKAQR--- | DKNARELKLANAAITDMQMRQ    | RDVAALDAKYTK 70 |                        |
|    |                       | ::: :         :       :     |                          |                 |                        |
| QY | 155                   | RYLDQFVPEKM-----            | RRELN-KAYCGFLRPGVSSSENLS | AVATGNWGGC      | AFGGDARL 205           |
|    |                       | :       :       :           |                          |                 |                        |
| Db | 71                    | ELADAKAENDALRDDVAAGRRRLHIK  | AVCQSVREATTASGV          | DNAAS-----      | PRL 119                |
| QY | 206                   | KALIQILAAAAAERDVVYFTFGDSEL  | MRDIYSMHI                | FLTERKLT        | TVGDVYKLLRYVNECR 265   |
|    |                       |                             |                          |                 |                        |
| Db | 120                   | -----ADTAERD--YFT-----      | -----LPERLITMOK          | OLEGTQK         | YVINEOCR 153           |
|    |                       |                             |                          |                 |                        |

```

RESULT 13
  ABP21
    endopeptidase (EC 3.4.-.-) - phage 21
    N:Alternate names: gene Rz protein; lysis protein Rz
    C:Species: phage 21
    C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 18-Jun-1999
    C:Accession: S22907

```

A;Molecule type: DNA  
A;Residues: 1-153 <BON>  
A;Cross-references: EMBL:M65239; NID:G215466; PIDN:AAA32351.1; PID:G215469  
C;Genetics:  
A;Gene: Rz  
C;Superfamily: phage PA2 endopeptidase  
C;Keywords: host cell lysis; hydrolase

|    |                       |   |                |            |             |  |
|----|-----------------------|---|----------------|------------|-------------|--|
|    | Query Match           | 5.4%;   | Score 86;      | DB 1;      | Length 153; |  |
|    | Best Local Similarity | 23.3%;  | Pred. No. 1.5; |            |             |  |
|    | Matches 42;           | Conservative 24;  | Mismatches 56; | Indels 58; | Gaps 8;     |  |
| QY | 104                   | CLIIITGEQYSEYTGAEYRWSRHEDGSDR-----DCERRCTEIVAILDHLFR        | 154            |            |             |  |
| Db | 14                    | CIIVCLSWAVNHRYDNAITYKAQR--DKNARELTLANRVITDIQMQRDVAAALDAKYTK | 70             |            |             |  |
| QY | 155                   | RYLDQFVPEKM-----RRELN-KAYCGFLRPGVSSENLSAVATGNWGCAGFGGDARL   | 205            |            |             |  |
| Db | 71                    | ELADAKAENDALRDDVAAGRRLHHKAVCQSVREATTASGVDNAAS-----PRL       | 119            |            |             |  |
| QY | 206                   | KALIQILAAAAAERDVVFTEFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRVNEECR | 265            |            |             |  |
| Db | 120                   | -----ADTAERD--YFT-----LRELITMQKOLEGTQKYNQECCR               | 153            |            |             |  |

RESULT 14  
S03572  
DNA-directed RNA polymerase (EC 2.7.7.6) chain B'' - Halobacterium salinarum  
C:Species: Halobacterium salinarum  
CDate: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 01-Dec-2000  
C:Accession: S03572  
R:Leffers, H.; Gropp, F.; Lottspeich, F.; Zillig, W.; Garrett, R.A.  
J. Mol. Biol. 206, 1-17, 1989  
A:Title: Sequence, organization, transcription and evolution of RNA polymerase subunit g  
A:Reference number: S03572; MUID:89199633; PMID:2495365  
A:Accession: S03572  
A:Molecule type: DNA  
A:Residues: 1-523 <LEF>  
A:Cross-references: EMBL:X57144; NID:g43538; PIDN:CAA40424.1; PID:g43540  
A:Note: the source is designated as Halobacterium halobium  
C:Superfamily: Halobacterium salinarum DNA-directed RNA polymerase chain B''  
C:Keywords: nucleotidyltransferase; transcription

| Query Match           | 5.3%;   | Score 85;              | DB 2;                  | Length 523; |
|-----------------------|---|------------------------|------------------------|-------------|
| Best Local Similarity | 22.3%;  | Pred. No. 9.4;         |                        |             |
| Matches 48;           | Conservative 29;  | Mismatches 70;         | Indels 68;             | Gaps 114    |
| QY 50                 | IEENGQGLQVDFANRFVGGVTSAGLVQ-----                              | EEIRFLINPELIISRLFTFEVL | DHN 102                |             |
| DB 209                | VERNREGLLVSPS--VSGSISFVTLVRALGLESDEEIVHRVSEDPEIVKFMLENLE--    | 264                    |                        |             |
| QY 103                | ECLIIITGTEQYSEY-----  | TGYAETYRWSRS-----      | HEDGSRDD-----          | 137         |
| DB 265                | EADVQTQEEAIEDLCQRVASGGQKNYQLKRANYVIDRYLLPHLHEDGVVEETRINKAYY   | 324                    |                        |             |
| QY 138                | CE--RRCTEIV-----  | AIDALHFR----           | RYLDQFVPEKMRRELNK----- | 171         |
| DB 325                | LCRMAEACFELALGRREADDDKHYANKRKLKVSGLMKDILFRTALNKLARDVKYQLERANM | 384                    |                        |             |
| QY 172                | -----AYCGFLRPGVSSSENLS-AVATGNWCGGAFG                          | 200                    |                        |             |
| DB 385                | RNRELTVNTVVRSDVLTLEHPIATGNWVGGRSG                             | 419                    |                        |             |

RESULT 15  
B84416  
DNA-directed RNA polymerase subunit B'' [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: B84416  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.;  
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, I.;  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: B84416  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-523 <STO>  
A:Cross-references: GB:AE004437; NID:g10582045; PIDN:AAG20694.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: rpoB''  
C:Superfamily: Halobacterium salinarum DNA-directed RNA polymerase chain B''

|    | Query Match   | 5.3%;            | Score 85;      | DB 2;      | Length 523; |
|----|---|------------------|----------------|------------|-------------|
|    | Best Local Similarity   | 22.3%;           | Pred. No. 9.4; |            |             |
|    | Matches 48;   | Conservative 29; | Mismatches 70; | Indels 68; | Gaps 11;    |
| Qy | 50 IEENGQGLQVDFANRFVGSGVTAGLVQ-----EEIRFLINPELIISRLFTFEVL DHN 102     | :                | :              | :          | :           |
| Dd | 209 VERNREGLLVEFSPS--VSGSISFVTLVRALGLESDDEIVHRVSEDP EIVKFMLENLE-- 264 | :                | :              | :          | :           |
| Qy | 103 ECLIIITGTEQYSEY----TGYAETYRWRS-----HEDGSRDD----- 137              | :                | :              | :          | :           |
| Dd | 265 EADVOTOEAEIEDIGORVASGGKNVOLKRANYVIDRYLLPHLHEDGVVEEETRINKAYV 324   | :                | :              | :          | :           |

QY 138 -CE--RRCTEIV-----AIDALHFR-----RYLDQFVPEKMRRELNK----- 171  
Db 325 LCRMAEACFELALGRREADDDKHAYANKRLKVSGLMKDLFRTALNKLARDVKYQLERANM 384  
QY 172 -----AYCGFLRPGVSSSENLS-AVATGNWGGGAFG 200  
Db 385 RNRELTNTVVRSDVLTTERLEHPIATGNWVGGRSG 419

Search completed: May 26, 2004, 18:47:32  
Job time : 11.2527 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:13:47 ; Search time 189.999 Seconds  
(without alignments)  
13517.524 Million cell updates/sec

Title: US-09-302-812-1  
Perfect score: 7242  
Sequence: 1 accggaagtgaacgaagcc.....aaattttcattacaaaaaa 4070

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/6333148/runat\_26052004\_150053\_5676/app\_query.fasta\_1.12437  
-DB=SPTREMBL\_25 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148 @CGN\_1\_1\_640 @runat\_26052004\_150053\_5676 -NCPUS=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL 25:

- 1: sp\_archaea:
- 2: sp\_bacteria:
- 3: sp\_fungi:
- 4: sp\_human:
- 5: sp\_invertebrate:
- 6: sp\_mammal:
- 7: sp\_mhc:
- 8: sp\_organelle:
- 9: sp\_phage:
- 10: sp\_plant:
- 11: sp\_rodent:
- 12: sp\_virus:
- 13: sp\_vertebrate:
- 14: sp\_unclassified:
- 15: sp\_rvirus:
- 16: sp\_bacteriap:
- 17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 5184  | 71.6        | 977    | 6  | 002776 bos taurus |

|    |        |      |      |    |        |                     |
|----|--------|------|------|----|--------|---------------------|
| 2  | 4694.5 | 64.8 | 976  | 4  | Q86W56 | Q86w56 homo sapien  |
| 3  | 4689.5 | 64.8 | 976  | 4  | Q7Z742 | Q7z742 homo sapien  |
| 4  | 4635.5 | 64.0 | 976  | 4  | Q9Y4W7 | Q9y4w7 homo sapien  |
| 5  | 4332.5 | 59.8 | 968  | 11 | Q88622 | Q88622 mus musculus |
| 6  | 4329   | 59.8 | 961  | 11 | Q80YQ6 | Q80yq6 mus musculus |
| 7  | 4295.5 | 59.3 | 972  | 11 | Q9QYM2 | Q9qym2 rattus norv  |
| 8  | 4125   | 57.0 | 920  | 11 | Q8CB72 | Q8cb72 mus musculus |
| 9  | 1063.5 | 14.7 | 768  | 5  | Q46043 | Q46043 drosophila   |
| 10 | 1061.5 | 14.7 | 723  | 5  | Q960N8 | Q960n8 drosophila   |
| 11 | 585    | 8.1  | 548  | 10 | Q9SKB3 | Q9skb3 arabidopsis  |
| 12 | 533    | 7.4  | 522  | 10 | Q8VYAL | Q8vya1 arabidopsis  |
| 13 | 481.5  | 6.6  | 781  | 5  | Q867X0 | Q867x0 caenorhabdi  |
| 14 | 475    | 6.6  | 764  | 5  | Q19637 | Q19637 caenorhabdi  |
| 15 | 472    | 6.5  | 485  | 5  | Q9N5L4 | Q9n5l4 caenorhabdi  |
| 16 | 335.5  | 4.6  | 368  | 5  | Q86GI4 | Q86gi4 toxoplasma   |
| 17 | 331    | 4.6  | 364  | 10 | Q9SKB4 | Q9skb4 arabidopsis  |
| 18 | 286.5  | 4.0  | 747  | 4  | Q86V50 | Q86v50 homo sapien  |
| 19 | 185.5  | 2.6  | 954  | 5  | Q9VU43 | Q9vu43 drosophila   |
| 20 | 185    | 2.6  | 508  | 12 | Q9E234 | Q9e234 helicoverpa  |
| 21 | 185    | 2.6  | 2081 | 10 | Q9LH98 | Q9lh98 arabidopsis  |
| 22 | 181.5  | 2.5  | 1618 | 16 | Q7UYM1 | Q7uym1 rhodopirell  |
| 23 | 181    | 2.5  | 571  | 5  | Q8MTN7 | Q8mtn7 trichinella  |
| 24 | 180    | 2.5  | 510  | 12 | Q99GU9 | Q99gu9 helicoverpa  |
| 25 | 179.5  | 2.5  | 1078 | 5  | Q963T1 | Q963t1 plasmodium   |
| 26 | 179    | 2.5  | 1160 | 11 | Q9ESC8 | Q9esc8 mus musculus |
| 27 | 177.5  | 2.5  | 1320 | 11 | Q7TPZ2 | Q7tpz2 mus musculus |
| 28 | 175.5  | 2.4  | 1099 | 10 | Q8W362 | Q8w362 oryza sativ  |
| 29 | 175.5  | 2.4  | 1099 | 10 | Q7XDA1 | Q7xda1 oryza sativ  |
| 30 | 175.5  | 2.4  | 1163 | 4  | Q9UHB7 | Q9uhb7 homo sapien  |
| 31 | 175.5  | 2.4  | 3942 | 11 | Q88737 | Q88737 mus musculus |
| 32 | 175.5  | 2.4  | 5327 | 5  | Q76891 | Q76891 drosophila   |
| 33 | 174.5  | 2.4  | 669  | 16 | Q7UE67 | Q7ue67 rhodopirell  |
| 34 | 174    | 2.4  | 3398 | 13 | Q8QFV6 | Q8qfv6 gallus gall  |
| 35 | 171.5  | 2.4  | 1444 | 5  | Q9VTN2 | Q9vtn2 drosophila   |
| 36 | 171.5  | 2.4  | 1514 | 5  | Q8SY55 | Q8sy55 drosophila   |
| 37 | 171.5  | 2.4  | 1616 | 4  | Q96PH3 | Q96ph3 homo sapien  |
| 38 | 171.5  | 2.4  | 1758 | 4  | Q7Z6E8 | Q7z6e8 homo sapien  |
| 39 | 171.5  | 2.4  | 1792 | 4  | Q7Z6E9 | Q7z6e9 homo sapien  |
| 40 | 171.5  | 2.4  | 5322 | 5  | Q9VPL9 | Q9vpl9 drosophila   |
| 41 | 171    | 2.4  | 791  | 13 | Q9DGL1 | Q9dgl1 fugu rubrip  |
| 42 | 171    | 2.4  | 1183 | 5  | Q9W3D1 | Q9w3d1 drosophila   |
| 43 | 170.5  | 2.4  | 1200 | 5  | Q25388 | Q25388 loligo peal  |
| 44 | 170.5  | 2.4  | 1226 | 16 | Q7UL12 | Q7ul12 rhodopirell  |
| 45 | 170.5  | 2.4  | 1576 | 3  | Q13308 | Q13308 candida alb  |

ALIGNMENTS

RESULT 1  
O02776  
ID O02776 PRELIMINARY; PRT; 977 AA.  
AC O02776;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN BPARG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97277328; PubMed=9115250;  
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase."  
RT J. Biol. Chem. 272:11895-11901(1997).  
DR EMBL; U78975; AAB53370.1; --  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.

KW Hydrolase.  
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 977  
Score: 5184.00 Matches: 977  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 71.58% Indels: 0  
DB: 6 Gaps: 0  
US-09-302-812-1 (1-4070) x 002776 (1-977)  
QY 258 ATGAGTGGCGCCCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317  
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
QY 318 ACTTCTCCGCGCCGCTCGGACGCCCGGAGCTTCCCGCGGAGGCGGCGCTCCTC 377  
Db 21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
QY 378 GATTCCAAAGGACGCTCCGGTGCAGTTTCAGGGTCCCGCGCTCCTCGTCAGGCTGGCCCTG 437  
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
QY 438 GGCCGGCGGAGACAGCACCTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAA 497  
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAA 557  
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLysHisSerLys 100  
QY 558 GAAACACACATACAGAGAGAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTAT 617  
Db 101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120  
QY 618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTGTATAGTACCAGTT 677  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 678 GAAAAAGGTACACAGTATTGTAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAA 737  
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 738 GAAGGGCCCACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 798 GAGCAGTTCAGTAATGCTAATGTCGATCGATCGTCCCAAGAGATGATCAGTGACACA 857  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
QY 858 AATAGTGAGGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAAGCTTGCGAATGCA 917  
Db 201 AsnSerGluGluSerArgAspAsnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 918 AAGCAGACGATGGAAGATGAACAGGCGCAGAGAGCCAGAGCCACCAAGAGTGTGGCAAG 977  
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 978 GCTTGGCATCCTGCAGAAGCCTGTGCAGGGTGTGACGAGGAGGAGACAGACGTGGTGTC 1037  
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260  
QY 1038 GAGAGCCCTTGTCCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097  
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
QY 1098 AGATTGAATAGACAAAGAAAGTAGTCTAGGAAATTCCTCCTCATTGTGAGAAAGAAAGTGA 1157  
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCAAAAATAGTTGTTCAGGATTCAGAAAGCAGAT 1217  
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
QY 1218 GAAGAGACAAAGTCCAGGTTTGTGATGAACAGGAAGATAGCAGTTCTGCTCAAAACAGCAA 1277  
Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAAGCGGTCTCTGCT 1337  
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
QY 1338 AAGGGAGGTGAGATTTCGATTACATTTCCAATTTGAAGGAGGAGAGTTCGAGCTGGAATG 1397  
Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGlyGlyGlyGlyMet 380  
QY 1398 AATGATGTAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGAPAGAGTGCAGAAAT 1457  
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAAATCACAGATCATTTTCATGAGAGTGCCCAA 1517  
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
QY 1518 GCAGAGGACAAAAGAAAGAACATGTGAAATGAAACATCAAGAGACAGAAAGGAGATC 1577  
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
QY 1578 CCTAATATACATCCACCTCACCTTCTCTCCAGATTAAGAAATGGCTTGGAACTCCTATTGAG 1637  
Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGGGATCCGGTGCCTCCCTGAGACCATCTGCCAAT 1697  
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
QY 1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAGTTCTTAAACCTTTCCCA 1757  
Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
QY 1758 ACACATTTTAAAGATTGTGGGACAAACAAAGCATGTTAAGATGCCTGTTCAGAACAAAC 1817  
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
QY 1818 TTGTACCTGTGGAAGATGAGAAATGGTGAGCGAGCTGCAGCAGCCGGTGGGAACTCATT 1877  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
QY 1878 CAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
QY 1938 TACATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTGATTGATTTCTGGGATAAG 1997  
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
QY 1998 GTACTAGAAGAGCAGAGCTCAACACTTGTATCATCTTCATCTTGCCTGATATGGTGAA 2057  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
QY 2058 ATTGCACCTGTCTGCCAAATATTTGTACCCAGCCAAATACCACTCCTGAAACAGAAAGATG 2117  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
QY 2118 AATCATTCACATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
QY 2178 TGACAGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGATTATCCAGATATT 2237  
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
QY 2238 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACAGAGAGCTTAAACCGCTC 2297

Db 661 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 680  
QY 2298 TTCTGCTACTTTAGAGAGTCAAGAGAAAAACCCACTGGGTGGTGACATTACAGA 2357  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACTCTGACTCGACTGCAT 2417  
Db 701 GlnSerLeuGluAspPheProGluTyrPheGluArgCysGluLysLeuLeuThrArgLeuHis 720  
QY 2418 GTCACCTACGAAGGTACCATAGAGGAAGAACGGCCAGGCGCATGTACAGGTGGATTGCA 2477  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2478 AACCGTTTCGTTGGAGTGGTGTAAACAGTGCAGGACTTGTCAAGAAATCCGCTTT 2537  
Db 741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
QY 2538 TTAATCAACCCCTGAGTTGATTGTTTACGGCTCTTCACTGAGGTGCTGGATCAATGAA 2597  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2598 TGTCTTATCATCACAGGTACTGACAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2658 CGTGGGCCCGGAGCCATGAAGACAGGAGCGAAGGAGCGACTGGCAGAGCGCGCACT 2717  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCACTTTGTGCCGAG 2777  
Db 821 GluIleValAlaIleAspAlaIleHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2778 AAGATCAGACGGGAGCTTAACAGGCTTACTGTGGATTCTTCGTCTCGTGGAGTTCTTCA 2837  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2838 GAGAACCTGTCTGAGTGGCTACAGGAACTGGGGCTGTGGTGCCTTTGGGGTGATGCT 2897  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
QY 2898 AGACTAAAGCCTTATACAGATCCTGGCAGCTGCTGAGCTGAGCGAGACGGTGTAT 2957  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 3018 GAGAGGAACTGACTGTGGAGAGTATATAGCTGCTGTACGATATTACATTAAGAA 3077  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
QY 3078 TGCAGAACTGCTCCACCCCGGACCAGACATCAAGCTTTATCCATTTCATATACCATGCA 3137  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 3138 GTTGAGTCTGTACAGACCAACCAACCGCCGGGACAAAGGACGGGGGCC 3188  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

RESULT 2

Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4694.50 Matches: 883  
Percent Similarity: 93.44% Conservative: 29  
Best Local Similarity: 90.47% Mismatches: 63  
Query Match: 64.82% Indels: 1  
DB: 4 Gaps: 1

US-09-302-812-1 (1-4070) x Q86W56 (1-976)

QY 258 ATGAGTGGGGCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCGCTGCA 317  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAlaThr 20  
QY 318 ACTTCTCCGCGCGCTCGGACGCCCGGAGCTTCCCGGCGGAGGCGGCGCTCCTC 377  
Db 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39  
QY 378 GATTCCAAGGACGCTCCGCTGCAGTTCAAGGTCGCGCGCTCCTCGTCAGGCTGCGCCCTG 437  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 438 GGCCGGGGGACAGCAGCAGGAGGCGGCCCTCTCTGTTTCAAAACAGAGACTATA 497  
Db 60 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTGAATCAGAAAAGTTTGATAGTAAA 557  
Db 80 ThrSerTyrMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 558 GAAACCAACAATACAGAGAGAAATCCATGATGAGTCTGTACAAAAGATAACTTTTAT 617  
Db 100 GluAsnAsnThrArgIleGluSerMetSerSerValGlnLysAspAsnPheTyr 119  
QY 618 CAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGGTTTGTATAGTACCAGTT 677  
Db 120 GlnHisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThr 139  
QY 678 GAAAAAGGTACACAGTATTGTAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAA 737  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 738 GAAGGGCCCACTCAGAACGGCTTTTGGAAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 798 GAGCAGTTCAATGCTAATGTGTCGATCAGTCGTCCTCCCAAGGATGATCAGTGACACA 857  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 858 AATAGTGAGGAGTAGAGATAATCAGCAGTTTGTGACATGTAAAGCTTCCGAATGCA 917  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 918 AAGCAGACGATGGAAGATGAACAGGCGCAGAGAAGCCAGAAAGCCAGCAAGTGTGGCAAG 977  
Db 220 LysGlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 978 GCTTGCCATCTCTCAGAGAGCCCTGTGCGAGGCTGTGAGGAGGAGGAGGAGGAGGAGG 1037



Db 240 SerCysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
 QY 1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAATGCGCAAC 1097  
 Db 260 GluSerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsn 279  
 QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTTCTCTCCATTGAGAAAGAAAGTGAA 1157  
 Db 280 LysLeuThrArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299  
 QY 1158 CCTGAGTCACCAATGGATGTAGATAAATTCCTGAGTGTGTGAGTGTGAGTGTGAGTGTGAGT 1217  
 Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
 QY 1218 GAAGAGACAAGTCCAGGTTTGTGATGAACAGGAAGATAGCAGTCTGCTCAAAACAGCAAT 1277  
 Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
 QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337  
 Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
 QY 1338 AAGGAGGTGAGATTGATTAATTCATTTCCAAATTTGAAGGAGGAGAGAGTGTGAGTGTGAGTGTGAGT 1397  
 Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
 QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGTGAGTGTGAGTGTGAGT 1457  
 Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerSerLeuAsnValGluCysArgAsn 399  
 QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTTAAATTCACAGATCATTTTCATGAGTGTGAGTGTGAGTGTGAGT 1517  
 Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisPheMetArgLeuProLys 419  
 QY 1518 GCAGAGGACAAAAGAAAGAAACAATGTGAATGAACATCAACATCAAGAAACAGAAAGAAAGTATC 1577  
 Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
 QY 1578 CCTAATATACATTCACCTCACCTTTCTCCAGATAAGAAATGGTTGGAACTCCTATTGAG 1637  
 Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
 QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGATCCGGCTGCCTCCCTTGAGACCATGTGCCAAT 1697  
 Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
 QY 1698 CACACAGTACTATTCGGGTAGATCTTTTGGCAATAGGAGAAAGTTCTTAAACCTTTCCCA 1757  
 Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
 QY 1758 ACACATTTTAAAGATTTGTGGACAACAAGCATGTTAAGATGCCTTGTTCAGAAACAAC 1817  
 Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
 QY 1818 TTGTACCTGTGGAAGATGAGAAATGGTGAAGCGAGCTGCAGGCAGCCGGTGGAACTCATT 1877  
 Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
 QY 1878 CAGACTGCATTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
 Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
 QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTCTGGGATAAG 1997  
 Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
 QY 1998 GTACTAGAAGAGCAGAAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGTTGAA 2057  
 Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
 QY 2058 ATTGCACTCTGTGCTGCAAAATATTTGTACCAGCAATATCCACTCTCTGAAACAGAAATG 2117

Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
 QY 2118 AATCATTTCCATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177  
 Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
 QY 2178 TGCACGTTTCCACGACCAATGCCAAGATGAATCAGAGTATTCAGATTATCCAGATTATCCAGATTAT 2237  
 Db 640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
 QY 2238 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACCCAGAGAAAGCTTAAACGCTC 2297  
 Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
 QY 2298 TTCTGCTACTTTAGAAAGAGTCCACAGAGAAACCCACTGGGTTGTTGACATTCACAAGA 2357  
 Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
 QY 2358 CAGAGTCTTGAAGATTTCCAGAGTGGGAAGATGTGAAAAAACTCCTGACTCGACTGCAT 2417  
 Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
 QY 2418 GTCACCTTACGAAGTACCATAGAAAGAAACCGCCAGGCGCATGTACAGTGTGATTTTGCAT 2477  
 Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
 QY 2478 AACCGTTTCTGTTGGAGGTGTGTAACCAAGTGCAGGACTTGTGCAAGAAAGAAATCCGCTTT 2537  
 Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759  
 QY 2538 TTAATCAACCCCTGAGTTGATTGTTTACGGCTCTTCACTGAGGTGTGATCACAATGAA 2597  
 Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
 QY 2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCGAAACATAC 2657  
 Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
 QY 2658 CGCTGGGCCCCGAGCCATGAAGACAGAGCGGAAAGGACGACTGGCAGAGCGGCACGACT 2717  
 Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspTrpGlnArgArgCysThr 819  
 QY 2718 GAGATCGTCCCATCGACCGCTCCACTTCCAGACGCTACCTCGACGAGTTTGTGCCCGAG 2777  
 Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
 QY 2778 AAGATCAGACGGGAGCTTAAACAAGGCTTACTGTGGATTTCTTCTGCTGGAGTTTCTTCA 2837  
 Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
 QY 2838 GAGAACCTGTCTGTCAGTGGCTACAGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897  
 Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
 QY 2898 AGACTAAAGCCTTAATACAGATCCTGGCAGCTGTGTAGCTGAGCGAGACGTTGTTTAT 2957  
 Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
 QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGACATTTTACAGCATGCATACATTCCTCACT 3017  
 Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
 QY 3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGTCTGTACGATATTACATTAAGAA 3077  
 Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
 QY 3078 TGCAGAAACTGCTCCACCCCGGACCAACATCAAGCTTTATCCATTCATATACCATGCA 3137  
 Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
 QY 3138 GTTGAGTCTGTACACAGACCAACCAACCGCCGGGACAAAGGACGGGG 3185  
 Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975



```
RESULT 3
Q7Z742 ID Q7Z742 PRELIMINARY; PRT; 976 AA.
AC Q7Z742;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052966; AAH52966.1; -.
KW Hypothetical protein.
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;

Alignment Scores:
Pred. No.: 0 Length: 976
Score: 4689.50 Matches: 882
Percent Similarity: 93.34% Conservative: 29
Best Local Similarity: 90.37% Mismatches: 64
Query Match: 64.75% Indels: 1
DB: 4 Gaps: 1

US-09-302-812-1 (1-4070) x Q7Z742 (1-976)
QY 258 ATGAGTGGGGCCCGGCTGTGAGCCCTTGACCAAGCGACCCCGTGGGACGCCGCTGCA 317
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAlaThr 20
QY 318 ACTTCTCCGGCCGCGCTCGGACGCCGAGCTTCCCGGCGAGGAGCGCGCTCCTC 377
Db 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39
QY 378 GATTCCCAAGGACGCTCCGGTGCAGTTTCAGGGTCCCGCGCTCCTCGTCAGGCTGCGCCCTG 437
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59
QY 438 GGCCGGGGGACAGCACCGGAGGCGGCCACCTCTCTTTGTTTCAAAACAGAAGACTATA 497
Db 60 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
```

```
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99
QY 558 GAAACCAACAATACAGAGAAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTAT 617
Db 100 GluAsnAsnAsnThrArgIleGluSerMetMetSerValGlnLysAspAsnPheTyr 119
QY 618 CAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGGTTTGTGATGATCACCAGTT 677
Db 120 GlnHisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThr 139
QY 678 GAAAAGGTACACAGTATTGGAAGCAGCATCAGACTGCGGCTATGTGTAAAGTGCAGAAAT 737
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159
QY 738 GAAGGGCCCACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179
QY 798 GAGCAGTTCAGTAATGTAATGTCGATCAGTCGTCCCAAGAGTATGATCAGTGCACACA 857
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199
QY 858 AATAGTGAGGAGTAGAGATAATCAGCAGTTTTGTGACACATGTAAGCTTGCAGTATGCA 917
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219
QY 918 AAGCAGACGATGGAAGATGAACAGGCGCAGAGAAGCCAGAACCCAGAGTGTGGCAAG 977
Db 220 LysGlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLys 239
QY 978 GCTTGCCATCCTGCAGAACGCTGTGCGGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1037
Db 240 SerCysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluLeuAspValValPro 259
QY 1038 GAGAGCCCTTGTGCGACACTGGCTGTGAGGATGTTGGTACTGGAGTGAATAATGCCAAC 1097
Db 260 GluSerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsn 279
QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTTCTCTCCATTTGAGAAAAGAGTGA 1157
Db 280 LysLeuThrArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299
QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCAAAATAATAGTTGTGAGATTTCAGAGCAGAT 1217
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319
QY 1218 GAAGAGACAAGTCCAGGTTTGTGATGAACAGAGAAGTAGCAGTTCTGCTCAACAGCAAT 1277
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339
QY 1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359
QY 1338 AAGGAGGTGAGATTGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTGGAGTGAATG 1397
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCTGAATGTAGAGTGCAGAAAT 1457
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399
QY 1458 TCTAAGCAACATGGGAGAAAGATTTCTAAAATCACAGATCATTTTCATGAGAGTGCCTAAA 1517
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisPheMetArgLeuProLys 419
QY 1518 GCAGAGGACAAAAGAAAAGAACAAATGTGAAATGAAACATCAAAAGACAGAAAGGAAGATC 1577
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439
QY 1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
```

Db 440 ProLysTyrValProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTCCCTCCCTTGAGACCATCTGCCAAT 1697  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1698 CACACAGTGAATTCGGGTAGATCTTTTCGGAATAGGAGAAGTTCCTAAACCTTCCCA 1757  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCCTTGTTCAGAACAAAC 1817  
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1818 TTGTACCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGAACTCATT 1877  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1878 CAGACTGCATCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997  
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1998 GTACTAGAAGAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGTGCTGATATGGTAAA 2057  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACCTCTGAAACAGAAGATG 2117  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 2118 AATCATTTCCATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGTAATGCTTTCTTC 2177  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGATTATCCAGATATT 2237  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 2238 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACCAGAGAGCTTAAACGCTC 2297  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2298 TTCTGCTACTTTAGAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTTCACAAGA 2357  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2358 CAGAGCTTTGAAGATTTTCCAGAGTGGGAAGATGTGAAAAACTCTGACTCGACTGCAT 2417  
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2418 GTCACCTTACGAAGGTACCATAGAGGAACCGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477  
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2478 AACCGTTTCGTTGGAGGTGGTGTAAACCAGTCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537  
Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759  
QY 2538 TTAATCAACCTTGAGTTGATTGTTTTCACGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2598 TGTCTTATCATCAGAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAACATAC 2657  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2658 CGCTGGGCCCGAGCCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGCGCGCAGCT 2717  
Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspTrpGlnArgHisCysThr 819

QY 2718 GAGATCGTCGCCATCGACCGCCCTCCACTTCAGACGCTACCTCGACAGATTGTGCCCGAG 2777  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCCTCTGGAGTTTCTTCA 2837  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTCTTTGGGGTGATGCT 2897  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 899  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 3078 TGCAGAAACTGCTCCACCCCGGACCAGACATCAAGCTTTATCCATTTCATATACCATGCA 3137  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 3138 GTTGAAGTCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975

RESULT 4

Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
AC Q9Y4W7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99380098; PubMed=10449915;  
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
RT hybridization.";  
RL Cytogenet. Cell Genet. 85:269-270(1999).  
DR EMBL; AF005043; AAB61614.1; -.  
DR Genew; HGNC:8605; PARG.  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;

Alignment Scores:

Pred. No.: 0 Length: 976  
Score: 4635.50 Matches: 874  
Percent Similarity: 92.83% Conservative: 32  
Best Local Similarity: 89.55% Mismatches: 69  
Query Match: 64.01% Indels: 1  
DB: 4 Gaps: 1

US-09-302-812-1 (1-4070) x Q9Y4W7 (1-976)

QY 258 ATGAGTGGCGCGCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGACCGCGCTGCA 317  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
QY 318 ACTTCTCCCGCGCGCTCGGACCGCGCGAGCTTCCCGCGCAGGAGCGCGCTCCTC 377  
Db 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGCTCCCGCGCTCCTCGTCAGGCTGGCCCTG 437  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 438 GCGCGCGCGGACAGCACCGGAGCGCGCCACCTCTCTTGTGTTTCAAACAGAGACTATA 497  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557  
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 558 GAAAAACAATAACAAGAGAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTAT 617  
Db 100 GluAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
QY 618 CAACATAACATGGAAAAATAGAAAAATGTTTCTCAGTAGGTTTTGATAAGTCACAGTT 677  
Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 678 GAAAAAGGTACACAGTATTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 798 GAGCAGTTGAGTAATGCTAATGTCGATCAGTCGTCCCAAGGATGATCACAGTCACACA 857  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAGCTTGGCAATGCA 917  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAGCCAGAGCCACCAGAGTGTGGCAAG 977  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 978 GCTTGCCATCCTGCAAGCCTGTGACGGGTGTACAGGAGGAGGAGACAGACGTGGTGTC 1037  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro 259  
QY 1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTGGTACTGGACTGMAAATGCCAAC 1097  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 1098 AGATTGAATAGACAAAGAAAGTAGTCTAGGAAATTTCTCCTCCATTGTAGAAAGAGTAA 1157  
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAATAAGTTGTGTCAGGATTCAGAGCAGAT 1217  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 1218 GAAGAGACAAGTCCAGTTTGTATGAACAGGAAGATAGCAGTTCTGTCTCAACAGCAAT 1277  
Db 320 GluGluThrSerProGlyPheAspGluGlnGlnAspGlySerSerSerGlnThrAlaAsn 339  
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCTCTGCT 1337  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1338 AAGGGAGGTGAGATTCCGATTACATTTCCAAATTTGAAGGAGGAGAGTCGAGCTGGAATG 1397

Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTAAATACACAGATCATTTTCATGAGAGTCCCAAA 1517  
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1518 GCAGAGGACAAAAGAAAGAAACAATGTGAAATGAAACATCAAGAAACAGAAAGGAGATC 1577  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1578 CCTAAATACATTCACCTCACCTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGATCCGGCTGCTCCCTTGAGACCATCTGCCAAT 1697  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1698 CACACAGTGACTATTCCGGTAGATCTTTTGGCAATAGGAGAGTTCTCTAAACCTTTCCCA 1757  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1758 ACACATTTTAAAGATTGTGGGACAAACAGCATGTTAAGATGCCTTGTTCAGAACAAAC 1817  
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisVallysMetProCysSerGluGlnAsn 519  
QY 1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACACTCAT 1877  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1878 CAGACTGCACCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTCTGGGATAAG 1997  
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1998 GTACTAGAAGACAGAGACTCAACACTTGTATCAGTCCATCTTGCCCTGATATGGTGAAA 2057  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 2058 ATTGCACCTCTGTCTGCCAAATATTTGTATCCAGCCAAATACCACCTCCTGAAACAGAGATG 2117  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 2118 AATCATTCATCACAAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTCTTC 2177  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 2178 TGCACGTTTCCACGACGCAATGCCAAGATGGAATCAGAGTATTCAGTTATCCAGATATT 2237  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 2238 AACTTCAATCGGTTGTTGAAGGACGTTTCATCAAGGAAACCAGAGAGCTTAAACCGCTC 2297  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAACCCACTGGGTTGGTGTGACATTCCACAAGA 2357  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCTGACTCGACTGCAT 2417  
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2418 GTCACCTTACGAAGGTACCATAGAGGAACCGCCAGGGCATGCTACAGGTGGATTTTGCA 2477



Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2478 AACCGTTTCGTTGGAGGTGTGTAACAGTGCAGGACTTGTGCAAGAAATCCGCTTT 2537  
Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759  
QY 2538 TTAATCAACCTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2598 TGTCTTATCATCACAGTACTGACGAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2658 CGCTGGCCCGGAGCCATGAAGACAGGAGCGAAAGGAGGAGTGGCAGAGGCGCAGACT 2717  
Db 800 ArgTyrSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACAGTTTGTGCCCGAG 2777  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGANTTCTTCTGCTGGAGTTCTTCA 2837  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2838 GAGAACCTGTCTGAGTGGCTACAGGAACTGGGGCTGTGCTGCTTGGGGTGATGCT 2897  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2898 AGACTAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACCTGGTTTAT 2957  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
QY 2958 TTCACCTTTGGGAGTCTCAGAACTGATGAGAGACATTTACAGCATGCATATTCCTCACT 3017  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 3018 GAGAGGAAACTGACTGTTGGAGAAGTATATTAAGCTGTGCTAGCTAGCATATTAACAATGAAGAA 3077  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 3078 TGCAGAAACTGTCCACCCCGGACCAACCAACCGCGGACAAAGGACGGG 3185  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 3138 GTTGAGTCTGTACACAGACCAACCAACCGCGGACAAAGGACGGG 3185  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975  
RESULT 5  
O88622 PRELIMINARY; PRT; 968 AA.  
AC O88622;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99380098; PubMed=10449915;  
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
RT hybridization."  
RL Cytogenet. Cell Genet. 85:269-270 (1999).  
DR EMBL; AF079557; AAC28735.1; -.  
DR MGD; MGI:1347094; Parg.

DR GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1B007 CRC64;

Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 59.82% Indels: 9  
DB: 11 Gaps: 5

US-09-302-812-1 (1-4070) x O88622 (1-968)

QY 258 ATGAGTGGGGCCCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGAGCGCGTGA 317  
Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLys---AlaArgTyrGlyAlaAlaGly 19  
QY 318 ACTTCTCCGCGCGCTCGGACGCGCGGAGCTTCCCGGAGGCGAGGCGCGTCTC 377  
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
QY 378 GATTCCAAAGGACGCTCCGCTGTCAGTTTCAGGTCCTCCGCTCCTCGTCAGGCTGCGCCCTG 437  
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
QY 438 GGCCGGCGGACAGCACCGGAGGCGCGCCACCTCTCTTGTGTTTCAAACAGAACTATA 497  
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTGAATCAGAAAGTTTCATAGTAAA 557  
Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
QY 558 GAAACAAACAATACAAGAGAAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTAT 617  
Db 97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116  
QY 618 CAACATAACATGGAAAATATAGAAATGTTTCTCAGCTAGGTTTGTATAGTCACCACTT 677  
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
QY 678 GAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGGTATGTGTAAAGTGGCAAT 737  
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
QY 738 GAAGGCGCACACTCAGAACGGCTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
QY 798 GAGCAGTTCAGTAATGCTAATGCTGATCAGTCGTCCCAAGAGGATGATCAGTGCACACA 857  
Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
QY 858 AATAGTGAAGAGAGTAGAGATAATCAGCAGTCTTTTGACACATGTAAAGCTTGCAGATGCA 917  
Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
QY 918 AAGCAGACGATGAAGATGAACAGGGCAGAGAAAGCCAGAGCCAGAGAGTGTGGCAAG 977  
Db 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233  
QY 978 GCTTGCCATCTGCAGAAAGCTGTGAGGGTGTGAGGAGGAGAGAGAGTGTGCTGCC 1037  
Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 253  
QY 1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAATGCAAC 1097  
Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATCTCTCCATTGAGAAAGAAAGTGAA 1157



|    |      |  |      |
|----|------|--|------|
| Db | 274  | LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProPheGluLysGluSerGlu      | 293  |
| QY | 1158 | CCTGAGTCACCAATGGATGTAGATAATTCAAAAAATAGTTGTCCAGGATTCAGAAAGCAGAT | 1217 |
| Db | 294  | ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp   | 313  |
| QY | 1218 | GAAGAGACAAAGTCAGGTTTGTATGAACACGAAAGATAGCAGTTCTGCTCAAACAGCAAAT  | 1277 |
| Db | 314  | GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn   | 332  |
| QY | 1278 | AAACCTTCAAGGTTCCAACCAAGAGAGCTGACACTGAGTTTGAGGAAGCGGTCTCTGTCT   | 1337 |
| Db | 333  | LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr   | 352  |
| QY | 1338 | AAGGAGGTGAGATTTCGATTACATTTCCAAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG  | 1397 |
| Db | 353  | LysGlySerGluValArgLeuHisPheGlnPheGlu--GlyGluAsnAsnAlaGlyThr    | 371  |
| QY | 1398 | AATGATGTGAATGCCAAACGACCTCGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT   | 1457 |
| Db | 372  | SerAspLeuAsnAlaLysProSerGlyAsnSerSerLeuAsnValGluCysArgSer      | 391  |
| QY | 1458 | TCTAAGCAACATGGGAGAAAGGATTTCTAAATTCACAGATCATTTTCATGAGAGTGCCCAA  | 1517 |
| Db | 392  | SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys   | 411  |
| QY | 1518 | GCAGAGGACAAAAGAAAAGAACAAATGTGAATGAACATCAAAGAACAGAAAGGAAGATC    | 1577 |
| Db | 412  | SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle   | 431  |
| QY | 1578 | CCTAAATACATTCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCTATTGAG     | 1637 |
| Db | 432  | ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu   | 451  |
| QY | 1638 | GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT   | 1697 |
| Db | 452  | GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer   | 471  |
| QY | 1698 | CACACAGTGACTATTCCGGTAGATCTTTTGCGAATAGGAGAAGTTCTCTAAACCTTTCCCA  | 1757 |
| Db | 472  | HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro   | 491  |
| QY | 1758 | ACACATTTTAAAGATTTGTGGGACAAACAAGCATGTTAAGATGCCTTGTTCAGAAACAAAAC | 1817 |
| Db | 492  | ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn   | 511  |
| QY | 1818 | TTGTACCCTGTGGAGATCAGAATGGTGAGCGGCTGCAGGCAGCCGGTGGGAACCTCAT     | 1877 |
| Db | 512  | LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle   | 531  |
| QY | 1878 | CAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG  | 1937 |
| Db | 532  | GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys   | 551  |
| QY | 1938 | TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTCTGGGATAAG    | 1997 |
| Db | 552  | TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys   | 571  |
| QY | 1998 | GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA   | 2057 |
| Db | 572  | ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys   | 591  |
| QY | 2058 | ATTGCACCTCTGTCTGCCAAATATTGTACCCAGCCAATACCACCTCTGAAACAGAAAGATG  | 2117 |
| Db | 592  | IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet   | 611  |
| QY | 2118 | AATCATTCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATAGCTTTCTTC   | 2177 |
| Db | 612  | AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe   | 631  |
| QY | 2178 | TGCACGTTTCCACGACGAATGCCAAGATGAAATCAGAGTATTCAGTTATCCAGATATT     | 2237 |

|    |      |  |      |
|----|------|--|------|
| Db | 632  | CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle     | 651  |
| QY | 2238 | AAC TTCATTCGGTTGTTTGAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAAACGCTC    | 2297 |
| Dd | 652  | AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu     | 671  |
| QY | 2298 | TTCTGCTACTTTAGAAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATT CACAAGA   | 2357 |
| Dd | 672  | PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg     | 691  |
| QY | 2358 | CAGAGTC TTGAAGATTTTCCAGAGTGGGAAAAGATGTGAAAAA CTCTGACTCGACTGCAT   | 2417 |
| Dd | 692  | GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis     | 711  |
| QY | 2418 | GTCAC TTTACGAAGGTACCATAGAAAGGAAACGCCACAGGCGCATGCTACAGTGGATTTTGCA | 2477 |
| Dd | 712  | ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla     | 731  |
| QY | 2478 | AACCGTTTCGTTGGAGGTGGTGTAACCACTGCAGGACTTGTGCAAGAGAAATCCGCTTT      | 2537 |
| Dd | 732  | AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluLeuArgPhe           | 751  |
| QY | 2538 | TTAATCAACCCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAAATGAA   | 2597 |
| Dd | 752  | LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu     | 771  |
| QY | 2598 | TGCTTTATCATCAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCGGAAACATAC       | 2657 |
| Dd | 772  | CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr     | 791  |
| QY | 2658 | CGCTGGCCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGGCGCACGACT    | 2717 |
| Dd | 792  | ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr     | 811  |
| QY | 2718 | GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG     | 2777 |
| Dd | 812  | GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu     | 831  |
| QY | 2778 | AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTCTTCTGTCCTGGAGTTTCTTCA     | 2837 |
| Dd | 832  | LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer     | 851  |
| QY | 2838 | GAGAACCTGTCTGAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT      | 2897 |
| Dd | 852  | GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla     | 871  |
| QY | 2898 | AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGTGCTGTAGCTGAGCGAGACGTGGTTTAT   | 2957 |
| Dd | 872  | ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr     | 891  |
| QY | 2958 | TTACCTTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT      | 3017 |
| Dd | 892  | PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr     | 911  |
| QY | 3018 | GAGAGGAAACTGACTGTTGGAGAAAGTATATAAGCTGCTGCTACGATATTACAAATGAAGAA   | 3077 |
| Dd | 912  | GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu     | 931  |
| QY | 3078 | TGCAGAAACTGCTCCACCCC CGGACCAGACATCAAGCTTTATCCATTTCATATACCATGCA   | 3137 |
| Dd | 932  | CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla     | 951  |
| QY | 3138 | GTTGAGTCCTGTACACAGACCACCAACCGCCGGGACAAAGGACGGGG 3185             |      |
| Dd | 952  | ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967             |      |

## RESULT 6

Q80YQ6  
REF001

YD ID

AC

DT

DT DT

DT DT

DE Poly (ADP-ribose) glycohydrolase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050892; AAH50892.1; -  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

Alignment Scores:  
Pred. No.: 0 Length: 961  
Score: 4329.00 Matches: 818  
Percent Similarity: 90.09% Conservative: 55  
Best Local Similarity: 84.42% Mismatches: 88  
Query Match: 59.78% Indels: 8  
DB: 11 Gaps: 4

US-09-302-812-1 (1-4070) x Q80YQ6 (1-961)

|    |     |   |     |
|----|-----|---|-----|
| QY | 258 | ATGAGTCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGACGCGCTGCA       | 317 |
| Db | 1   | MetSerAlaGlyProGlyTyrGluProCysThrLysArgProArgTyrGlyAlaAlaGly    | 20  |
| QY | 318 | ACTTCTCCGCCCGCCGCTCGGACGCCCGGAGCTTCCCGGCGAGCGGCGGCGCTCCTC       | 377 |
| Db | 21  | ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu       | 40  |
| QY | 378 | GATTCCAGGACGCTCCGGTGCAGTTCAGGTCCCGCGCTCCTCGTCAGGCTGCGCCCTG      | 437 |
| Db | 41  | AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer    | 60  |
| QY | 438 | GGCCGGCGGGACAGCACCGAGGCGCGCCACCTCTCTTGTGTTTCAACACAGAAGACTATA    | 497 |
| Db | 61  | GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle    | 80  |
| QY | 498 | ACCAGTTGGATGGACACTAAAGGAATCAAGCACAGTTGAATCAGAAAGTTTGCATAGTAAA   | 557 |
| Db | 81  | ThrThrTyrMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys        | 97  |
| QY | 558 | GAACAACAAATAACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAGATAACTTTTAT     | 617 |
| Db | 98  | GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr    | 117 |
| QY | 618 | CAACATAACATGGAAAAAATTAGAAAATGTTTCTCAGCTAGGTTTGTGATAAGTCACCAGTT  | 677 |
| Db | 118 | ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr    | 137 |
| QY | 678 | GAAAAAGGTACACAGTATTGTAAGCAGCATCAGACTCGCGCTATGTGTAGTGGCAGAAAT    | 737 |
| Db | 138 | GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysLysTrpGlnAsn | 157 |
| QY | 738 | GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA    | 797 |
| Db | 158 | GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro    | 177 |
| QY | 798 | GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCTCCCAAGGATGATCACAGTGACACA   | 857 |
| Db | 178 | LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr    | 197 |
| QY | 858 | AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTTTGACACATGTAAAGCTTGCAGATGCA  | 917 |
| Db | 198 | AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr    | 217 |
| QY | 918 | AAGCAGACGATGGAAGATGAACAGGGCAGAGAGCCAGAGCCAGAGCTGTGGCAAG         | 977 |

QY 2058 ATTGCACCTCTGTGTCACCAATATTTGTACCCAGCCCAATACCACCTCTCTGAAACAGAGATG 2117  
Db |||||||  
593 IleAlaLeuCyLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 612  
QY 2118 AATCATTCATCAATGTACAGGAACAGATTGCCAGTCTTTTAGCTTAATGCTTTCTTC 2177  
Db |||||||  
613 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 632  
QY 2178 TGCACGTTTCCACGCGCAATGCCAAGATGAATCAGAGTATTCAGTTATCCAGATATT 2237  
Db |||||||  
633 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 652  
QY 2238 AACTTCAATCGTTGTTGAAGGACGTTTCATCAAGGAAACACAGAGAAAGCTTAAACGCTC 2297  
Db |||||||  
653 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 672  
QY 2298 TTCTGCTACTTTAGAGAGTCACAGAGAAAAACCCACTGGTTGGTGACATTCACAAGA 2357  
Db |||||||  
673 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 692  
QY 2358 CAGAGCTTTGAAGATTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417  
Db |||||||  
693 GlnSerLeuGluAspPheProGluTyrPheGluArgCysGluLysProLeuThrArgLeuHis 712  
QY 2418 GTCACTTACGAAGGTACATAGAAGGAAACCGCCAGGCATGCTACAGGTGATTTGCA 2477  
Db |||||||  
713 ValThrTyrGluGlyThrIleGluGlyArgGlyMetLeuGlnValAspPheAla 732  
QY 2478 AACCGTTTTCGTTGGAGTGGTGTAAACAGTGCAGGACTGTGCAAGAAGAAATCCGCTTT 2537  
Db |||||||  
733 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 752  
QY 2538 TTAATCAACCCCTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db |||||||  
753 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 772  
QY 2598 TGTCTTATCATCACAGGTACTGACGACTACAGTGAATACACAGGCTATGCCGAACATAC 2657  
Db |||||||  
773 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 792  
QY 2658 CGCTGGGCCCGGAGCCATGAACAGAGGAGGAAAGGGACGACTGGCAGAGCGCCGACT 2717  
Db |||||||  
793 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 812  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACTCGACCAAGTTTGTGCCGAG 2777  
Db |||||||  
813 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 832  
QY 2778 AAGATCAGACGGGAGCTTAAACAGGCTTACTGTGGATTCTTCGTCCTGGAGTTTCTTCA 2837  
Db |||||||  
833 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 852  
QY 2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGCTGTGGTGCCTTTGGGGTGATGCT 2897  
Db |||||||  
853 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 872  
QY 2898 AGACTAAAAGCCTTAATACAGATCTCGGAGCTGTGTAGCTGAGCGAGACGTGTTTAT 2957  
Db |||||||  
873 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 892  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTTACAGCATGCATACATTCCTCACT 3017  
Db |||||||  
893 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 912  
QY 3018 GAGAGGAACCTGACTGTTGGAGAGATATATAAGCTGCTGTACGATATTACATGAAGAA 3077  
Db |||||||  
913 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 932  
QY 3078 TGCAGAACTGCTCCACCCCGGACCAACATCAAGCTTTTATCCATTATATATACCATGCA 3137  
Db |||||||  
933 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 952

QY 3138 GTTGAGTCTCTGTACACAGACCACCAAC 3164  
Db |||||||  
953 ValGluSerSerAlaGluThrThrAsp 961  
RESULT 7  
Q9QYM2 PRELIMINARY; PRT; 972 AA.  
AC Q9QYM2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN PARG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BUF; TISSUE=Colon;  
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,  
RA Aoki Y., Nakgama H., Sugimura T.;  
RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase  
RT (Parg).";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB019366; BAA87901.1; --  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 972  
Score: 4295.50 Matches: 808  
Percent Similarity: 88.75% Conservative: 60  
Best Local Similarity: 82.62% Mismatches: 101  
Query Match: 59.31% Indels: 9  
DB: 11 Gaps: 4  
US-09-302-812-1 (1-4070) x Q9QYM2 (1-972)  
QY 258 ATGAGTGGGGCCCCGGCTGTGAGCCCTGCACCACGACCCCGCTGGGACGCCGCTGCA 317  
Db |||||||  
1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20  
QY 318 ACTTCTCCGCGCGCTCGGACGCCCGGAGCTTCCCGGCGAGGCGGCGCTCCTC 377  
Db |||||||  
21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnLysArgValLeu 40  
QY 378 GATTCCAAGGACGCTCCGGTGCAGTTTCAGGGTCCCGCGCTCCTCGTCAGGCTGCGCCCTG 437  
Db |||||||  
41 AspProLysAspAlaProValGlnPheArgValProProSerSerAlaCysValSer 60  
QY 438 GGCCTGGGCGGACAGCACCGGAGGCGGCCACCTCTCTGTTTCAAAACAGAAGACTATA 497  
Db |||||||  
61 GlyArgAlaGlyProHisArgGlySerValThrSerPheValPheLysGlnLysProIle 80  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAAGTTTGCATAGTAAA 557  
Db |||||||  
81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 97  
QY 558 GAAACAAACAATACAGAGAAGAATCCATGATGATGTTCTGTGTACAAAAGATAACTTTTAT 617  
Db |||||||  
98 GluAsnAsnAsnThrArgThrAspProMetMetSerSerValGlnLysAspAsnPheTyr 117  
QY 618 CAACATAACATGGAAAATAGAAAATGTTTCTCAGCTAGGTTTGTATAGTCACCACTT 677  
Db |||||||  
118 ProHisLysValGluLysLeuGlyAsnValProGlnLeuAsnLeuAspLysSerProThr 137  
QY 678 GAAAAGGTACACAGTATTGAAAGCAGCATCAGACTGGGCTATGTGTAAGTGGCAGAA 737  
Db |||||||  
138 GluLysSerThrProTyrLeuAsnGlnGlnGlnThrAlaGlyValCysLysTrpHisSer 157



QY 738 GAAGGGCCCACTCAGAACCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 158 AlaGlyGluArgAlaGluGlnLeuSerAlaSerGluProSerAlaValThrGlnAlaPro 177  
QY 798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCAGTCCGCCAAAGGATGATCACAGTGACACA 857  
Db 178 LysGlnLeuSerAsnAlaAsnIleAspGlnSerProProThrAspGlyHisSerAspThr 197  
QY 858 AATAGTGAGGACAGTAGAGATAATCAGCAGTTTTGCACATGTAAAGCTTCCGAATGCA 917  
Db 198 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProValLysLeuAlaAsnAla 217  
QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAGCCACCAGAAGTGTGGCAAG 977  
Db 218 LysGlnThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerAla 234  
QY 978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCCAGGGTGTCCAGGAGGAGACAGCTGGTGTCC 1037  
Db 235 SerCysGlnCysGlyGlnAspCysAlaGlyCysGlnArgGluGluAlaAspValIlePro 254  
QY 1038 GAGAGCCCTTGTCCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097  
Db 255 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlySerLysAsnAspAsn 274  
QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATCTCCTCCATTTGAGAAAGAAAGTGAA 1157  
Db 275 LysLeuThrGlyGlnGluSerGlyLeuGlyAspSerProProPheGluLysGluSerGlu 294  
QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217  
Db 295 ProGluSerProMetAspValAspAsnSerLysThrSerCysGlnAspSerGluAlaAsp 314  
QY 1218 GAAGAGACAAAGTCCAGGTTTGTATGAACAGGAAGATAGCAGT-----TCTGCTCAAACA 1271  
Db 315 GluGluAlaSerProValPheAspGluGlnAspAspGlnAspAspArgSerSerGlnThr 334  
QY 1272 GCAAAATAACCTTCAAGTTCACCAACCAAGAGAAAGCTGACACTGAGTTGAGGAAGCGGTCC 1331  
Db 335 AlaAsnLysLeuSerSerArgGlnAlaArgGluValAspGlyAspLeuArgLysArgTyr 354  
QY 1332 TCTGCTAAGGGAGGTGAGATTGATTCATTTCCAAATTTGAAGGAGGAGAGAGTCGAGCT 1391  
Db 355 LeuThrLysGlySerGluIleArgLeuHisPheGlnPheGluGlyGly---SerAsnAla 373  
QY 1392 GGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGC 1451  
Db 374 GlyThrSerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValAspGly 393  
QY 1452 AGAAATCTAAGCAACATGGAGAAAGGATTTCTAAATCACAGATCATTTTCATGAGAGTG 1511  
Db 394 ArgSerSerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheValArgIle 413  
QY 1512 CCCAAAGCAGAGGACAAAAGAAAGAAACAATGTGAATGAACATCAAAGAACAGAAAGG 1571  
Db 414 ProLysSerGluAspLysArgLysGluGlnCysGluValArgHisGlnArgAlaGluArg 433  
QY 1572 AAGATCCCTAAATACATTCACCTCACCTTCTCCAGATAAGAAATGGCTTGGAACTCCT 1631  
Db 434 LysIleProLysTyrValProProAsnLeuProProAspLysLysTrpLeuGlyThrPro 453  
QY 1632 ATTGAGGAGATGAGGAGAATGCCAAGGTGGGATCCGGCTGCCTCCCTTGAGACCATCT 1691  
Db 454 IleGluGluMetArgLysMetProArgCysGlyValArgLeuProLeuLeuArgProSer 473  
QY 1692 GCCAATCACACAGTACTATTCCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCT 1751  
Db 474 AlaSerHisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysPro 493  
QY 1752 TTCCCAACACATTTTAAAGATTTGTGGGACAAACAGCATGTTAAGATGCCTTGTTTCAGAA 1811  
Db 494 PheProThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGlu 513

QY 1812 CAAAACCTTGTAACCTGTGGAAGATGAGAAATGGTGAGCGAGCTGCAGCGCGGTGGGAA 1871  
Db 514 GlnAsnLeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGlu 533  
QY 1872 CTCATTGAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATT 1931  
Db 534 LeuIleGlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIle 553  
QY 1932 CTGAAGTACAATGTGGCATATTTCTAAGAAATGGGACTTTTACAGCTTTGATTGTTCTGG 1991  
Db 554 LeuLysTyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrp 573  
QY 1992 GATAAGGTACTAGAAGAGCAGAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATG 2051  
Db 574 AspLysValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMet 593  
QY 2052 GTGAAATGTCACCTCTGTCTGCCAAATATTTGTACCCAGCCCAATACCACTCCTGAAACAG 2111  
Db 594 ValLysIleAlaLeuCysLeuProAsnIleCysThrGlnProIleProleuLeuLysGln 613  
QY 2112 AAGATGAATCATTCATCACAAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAATGCT 2171  
Db 614 LysMetAsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAla 633  
QY 2172 TTCTTCTGCACGTTTCCACGACGCAATGCAAGATGAAATCAGAGTATTCAGTTATCCA 2231  
Db 634 PhePheCysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrPro 653  
QY 2232 GATATTAACTTCAATCGGTTGTTTGAAGACGCTTCATCAAGGAAACCCAGAGAAAGCTTAAA 2291  
Db 654 AspIleAsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLys 673  
QY 2292 ACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAACCCACTGGGTGGTGACATTC 2351  
Db 674 ThrLeuPheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPhe 693  
QY 2352 ACAAGACAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGA 2411  
Db 694 ThrArgGlnSerLeuGluAspPheProGluTrpGluArgCysAspLysProLeuThrArg 713  
QY 2412 CTGCATGTCACTTACGAAGGTACCATAGAAAGAAACGGCCAGGCGCATGTACAGGTGGAT 2471  
Db 714 LeuHisValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAsp 733  
QY 2472 TTTGCAAAACCGTTTTCGTGGAGGTGGTGTAAACAGTGCAGGACTTGTGTCAAGAAATC 2531  
Db 734 PheAlaAsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluIle 753  
QY 2532 CGCTTTTAAATCAACCTGAGTTGATTGTTTTCAGGCTCTTCACTGAGGTGCTGATCAC 2591  
Db 754 ArgPheLeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHis 773  
QY 2592 AATGAATGTCTTATCATCAGGTACTGAGCAGTACAGTGAATACACAGGTATGCGGAA 2651  
Db 774 AsnGluCysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGlu 793  
QY 2652 ACATACCGCTGGGCCCGAGCCATGAAGACAGGAGCGAAAGGACGACTGGCAGAGCGC 2711  
Db 794 ThrTyrArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgCys 813  
QY 2712 ACGACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCACTTTGTG 2771  
Db 814 CysThrGluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheVal 833  
QY 2772 CCGGAGAAGATCAGACGGGAGCTTAAACAGGCTTACTGTGGATTTCTTCGTCCTGGAGTT 2831  
Db 834 ProGluLysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyVal 853  
QY 2832 TCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGT 2891  
Db 854 ProProGluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGly 873  
QY 2892 GATGCTAGACTAAAGCCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTG 2951



Db 874 AspAlaArgLeuLysAlaLeuIleGlnLeuAlaAlaAlaAlaGluArgAspVal 893

QY 2952 GTTTATTTCACCTTTGGGACTCAGAACTGATGAGACATTTACAGATGCATACATTC 3011

Db 894 ValTyrPheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPhe 913

QY 3012 CTCACCTGAGAGGAACTGACTGTTGGAGAGATATATAAGCTGCTGCTACATATTACAAT 3071

Db 914 LeuThrGluArgLysLeuAsnValGlyLysValTyrArgLeuLeuLeuArgTyrTyrArg 933

QY 3072 GAAGAATGCAGAACTGCTCCACCCCGGACCAAGCTTTATCCATTCATATAC 3131

Db 934 GluGluCysArgAspCysSerProGlyProAspThrLysLeuTyrProPheIleTyr 953

QY 3132 CATGCAGTTGAGTCTCTGTACACAGACCAACCCAGCCGGGACAAAGGACGGGG 3185

Db 954 HisAlaAlaGluSerSerAlaGluThrSerAspGlnProGlyGlnArgThrGly 971

RESULT 8

Q8CB72 PRELIMINARY; PRT; 920 AA.

AC Q8CB72; (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Poly.

GN PARG.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Bone;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573 (2002).

DR EMBL; AK036656; BAC29519.1; --

DR MGD; MGI:1347094; Parg.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Alignment Scores:

Pred. No.: 1.77e-311 Length: 920

Score: 4125.00 Matches: 782

Percent Similarity: 89.87% Conservative: 52

Best Local Similarity: 84.27% Mismatches: 86

Query Match: 56.96% Indels: 8

DB: 11 Gaps: 4

US-09-302-812-1 (1-4070) x Q8CB72 (1-920)

QY 258 ATGAGTGGGSCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGACGCGCTGCA 317

Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLysArgProArgTyrGlyAlaAlaGly 20

QY 318 ACTTCTCCGCGCGCTCGGACGCCCGGAGCTTCCCGGAGGCGGCGCTCCTC 377

Db 21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 40

QY 378 GATTCCAGGACGCTCCGGTGCAGTTCCAGGTCGCGCGCTCCTCGTCAGGCTGCGCCCTG 437

Db 41 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60

QY 438 GGCCGGGCGGACAGCAGCCGAGGCGCCACCTCTCTTGTGTTTCAACACAGAACTATA 497

Db 61 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 80

QY 498 ACCAGTTGGATGGACACATAAAGGAATCAAGACAGATTGAATCAGAAAAGTTGCATAGTAAA 557

Db 81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 97

QY 558 GAAACAAACAATACAAGAGAAGATCCATGATGAGTTCTGTACAAAAGATAAAGTTTAT 617

Db 98 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 117

QY 618 CAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGGTTTGTGTAAGTCACCAGTT 677

Db 118 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 137

QY 678 GAAAAAGGTACACAGATATTGAAGCAGCATCAGACTGCGGCTATGTGTAACTGGCAGAA 737

Db 138 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 157

QY 738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGTAACTCTGGTACCA 797

Db 158 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 177

QY 798 GAGCAGTTTCAGTAATGCTAATGTCGATCAGTCCTCCCAAAGGATGATCAGTGACACA 857

Db 178 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 197

QY 858 AATAGTGAGGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAAGCTTGCGAATGCA 917

Db 198 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 217

QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAGCCAGAACCCAGAAAGTGTGCGCAAG 977

Db 218 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 234

QY 978 GCTTGCCATCCTGCAGAACGCTGTGCAGGGTGTACAGGAGGAGACAGACGCTGGTGCTCC 1037

Db 235 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 254

QY 1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTGTGCTGAGCTGAAATAATGCCAAC 1097

Db 255 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 274

QY 1098 AGATTGAATAGACAAGAACTAGTCTAGGAAATTCCTCCTCCATTGAGAAAAGAAAGTGA 1157

Db 275 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 294

QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCCTCCTCCTCAGGATTCAGAAAGCAGAT 1217

Db 295 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 314

QY 1218 GAAGAGACAAAGTCCAGGTTTTCATGAACAGGAAGATAGCAGTTCTGCTCAACAGCAAAT 1277

Db 315 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 333

QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAAGCGGCTCTGCT 1337

Db 334 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 353

QY 1338 AAGGAGGTGAGATTTCGATTACATTTTCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397

Db 354 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 372

QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCCTGAATGTAGAGTCAGAAAT 1457

Db 373 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 392

QY 1458 TCTAAGCAACATGGGAGAAAGGATTTCTAAATCAGATCATTTTCATGAGAGTGCCCAA 1517

Db 393 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 412

QY 1518 GCAGAGCAAAAGAAAGAAACAATGTGAATGAAACATCAAGAGACAGAAAGGAGATC 1577

Db 413 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 432

QY 1578 CCTAAATACATTCACCTCCACCTTCTCCAGATAAAGAAATGGCTTGGAACCTCCTATTGAG 1637

|||||ProLysTyrIleProProAsnLeuProProGluLysLysTyrLeuGlyThrProIleGlu 452  
1638 GAGATGAGGAGATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697  
453 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 472  
1698 CACACAGTGACTATTCGGGTAGATCTTTTGGCAATAGGAGAAAGTTCTTAAACCTTTCCCA 1757  
473 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 492  
1758 ACACATTTTAAAGATTGTGGGACAAACAGCATGTTAAGATGCCCTTGTTCAGAACAAAC 1817  
493 ThrHisTyrLysAspLeuTyrAspAsnLysHisValLysMetProCysSerGluGlnAsn 512  
1818 TTGTACCCCTGTGGAGATGAGAAATGGTGAGCGAGCTGCAGGCAGCCGGTGGAACTCAT 1877  
513 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTyrGluLeuIle 532  
1878 CAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
533 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 552  
1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTTGATTGATTTCTGGGATAAG 1997  
553 TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuValAspPheTyrAspLys 572  
1998 GTACTAGAAGAAGCAGAGCTCAACACTTGTTATCAGTCCATCTTGCTGATATGGTGAA 2057  
573 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 592  
2058 ATTGCACCTCTGTCTGCCAAATATTTGTACCCAGCCCAATACCACCTCTGAAACAGAGATG 2117  
593 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 612  
2118 AATCATTTCCATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTCTTC 2177  
613 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 632  
2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGTTATCCAGATATT 2237  
633 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 652  
2238 AACTTCAATCGGTTGTTTGAAGACGTTTCATCAAGGAAACAGAGAAGCTTAAACGCTC 2297  
653 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 672  
2298 TTCTGCTACTTTAGAGAGCTCACAGAGAAAAACCCACTGGGTGGTGACATTCACAAGA 2357  
673 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 692  
2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAAATCCTGACTCGACTGCAT 2417  
693 GlnSerLeuGluAspPheProGluTyrPheGluArgCysGluLysProLeuThrArgLeuHis 712  
2418 GTCACTTACGAAGGTACCATAGAGGAACGGCCAGGCGATGCTACAGGTGGATTTTGCA 2477  
713 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 732  
2478 AACCGTTTCGTTGGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537  
733 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 752  
2538 TTAATCAACCTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
753 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 772  
2598 TGTCTTATCATCAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657  
773 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 792  
2658 CGCTGGGCCCGAGCCCATGAAGACAGGACGGAAGGACGACTGGCAGAGCGCCAGCT 2717

793 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 812  
2718 GAGATCGTCGCCCATCGACGCCCTCCACTTCAGACGCTACCTCGACAGTTTGTGCCCGAG 2777  
813 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 832  
2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTCTTCGTCCTCGAGTTTCTTCA 2837  
833 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 852  
2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897  
853 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspVal 872  
2898 AGACTAAAGCCTTAATACAGATCCTGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957  
873 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 892  
2958 TTCACCTTTGGGACTCAGAACTGATGATGAGACATTTACAGCATGCATACATTCCTCACT 3017  
893 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMethisThrPheLeuThr 912  
3018 GAGAGGAACTGACTGTTGGAGAA 3041  
913 GluArgLysLeuAspValGlyGlu 920  
RESULT 9  
O46043 PRELIMINARY; PRT; 768 AA.  
ID O46043  
AC O46043;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE PARG protein.  
GN PARG OR EG:114E2.1 OR CG2864.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ame J.-C., Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding *Drosophila*  
RT poly(ADP-ribose) glycohydrolase.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 46-768 FROM N.A.  
RA Murphy L., Harris D., Barrell B.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 46-768 FROM N.A.  
RA Benos P.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003428; AAF45886.1; -.  
DR EMBL; AF079556; AAC28734.1; -.  
DR EMBL; Z98254; CAB10913.1; -.  
DR FlyBase; FBgn0023216; Parg.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;

Alignment Scores:  
Pred. NO.: 1.18e-73 Length: 768  
Score: 1063.50 Matches: 244  
Percent Similarity: 57.31% Conservative: 93  
Best Local Similarity: 41.50% Mismatches: 198  
Query Match: 14.69% Indels: 53  
DB: 5 Gaps: 15

US-09-302-812-1 (1-4070) x O46043 (1-768)

QY 1506 AGAGTCCCCAAGCAGGAGGACAAAGAAAGAAACAAATGTGAATGAACATCAAGAACA 1565  
Db 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGlu----- 62  
QY 1566 GAAAGGAAGATCCCTAAATACATTCACCTCACCT-----TCTCCAGATAAGAAATGG 1619  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTrp 74  
QY 1620 CTTGGAACCTCTATTGAGGAGATG-----AGGAGAATGCCAAGGTGTGGATCCGGCTG 1673  
Db 75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94  
QY 1674 CTTCCCTTGAGACCATTCTGCCAATCACACAGTACTATTCGGGTAGATCTTTGCGCAATA 1733  
Db 95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGln-----LeuProIle 111  
QY 1734 GGAGAAGTTCCTAAACCTTTCACACACATTTTAAAGATTGTGGGACAAACAGCATGTT 1793  
Db 112 ArgGluThrPro---ProArgProTyrLysSerProGlyLysTrpAspSerGluHisVal 130  
QY 1794 AAGATGCCTTGTTCAGAACAAACTTGTACCCCTGTGGAAGATGAGAATGGTGGCGAGCT 1853  
Db 131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150  
QY 1854 GCAGGACCGCGTGGGAACCTATTACAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAG 1913  
Db 151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170  
QY 1914 AACCTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1973

Db 171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190  
QY 1974 GCTTTGATTGATTCTCTGGGATAAGCTACTAGAAAGCAGAGAGCTCAACACTTGATCAG 2033  
Db 191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210  
QY 2034 TCCATCTTCCTGATATGGTGAATATTCACCTCTGTCTGCCAAATATTGTACCCAGCCA 2093  
Db 211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230  
QY 2094 ATACCACTCTGAAACAGAGATGAATCATTCATCCATCACAATGTCCAGGAACAGATTGCC 2153  
Db 231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSer 250  
QY 2154 AGTCTTTTAGTAATGCTTTCTTCTGCAGCTTTCCAGCAGCAATGCC---AAGATGAAA 2210  
Db 251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270  
QY 2211 TCAGAGTATTCAGATTATCCAGATATTAACTTCAATCGGTTGTTGAGGACGTTTCATCA 2270  
Db 271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290  
QY 2271 AGGAACCCAGAGAGCTTAAACCGCTCTTCTGTCTACTTTAGAAAGATC-----ACAGAG 2324  
Db 291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310  
QY 2325 AAAAA-----CCCACTGGGTTGGTGACATTCACAAGACAGAGT-----CTT 2366  
Db 311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330  
QY 2367 GAAGATTTTCAGAGTGGGAAAGATGTGAAAAAATCCTCTG-----ACTCGACTGCATGTC 2420  
Db 331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350  
QY 2421 ACTTACGAAGGTACCATAGAAGAAACGGCCAGGCGCATGTCTACAGGTGATTTGCAAAAC 2480  
Db 351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370  
QY 2481 CGTTTCGTTGGAGGTGGTGTAAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTAA 2540  
Db 371 LysTyrLeuGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390  
QY 2541 ATCAACCCCTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600  
Db 391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410  
QY 2601 CTTATCATCACAGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660  
Db 411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430  
QY 2661 TGGGCCCGGAGCCATGAAGACAGGAGCGAAGCGGACGACTGGCAGAGCGCGCAGCTGAG 2720  
Db 431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450  
QY 2721 ATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACAGTTTGTGCCCGAGAAG 2780  
Db 451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470  
QY 2781 ATCAGACGGGAGCTTAAACAGGCTTACTGTGGATTCTTCTCGT----- 2822  
Db 471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490  
QY 2823 CCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGCCTACAGGAACTGGGGCTGTGGTGCC 2882  
Db 491 ProGly-----ValAlaThrGlyAsnTrpGlyCysGlyAla 502  
QY 2883 TTTGGGGGTGATGTAGACTAAAGCCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAG 2942  
Db 503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522  
QY 2943 CGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGACACATTTACAGCATG 3002  
Db 523 ArgProLeuAlaTyrThrThrPheGlyAsnValGluPheArgAspPheHisGluMet 542



|   |   |   |              |
|---|---|---|--------------|
| QY  | 3003  | CATACATTCCTCACTGAGAGGAACTGACTGTTGGAGAGTATATATAAGCTGCTGCTACGA  | 3062         |
| Db  | 543   | TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg  | 561          |
| QY  | 3063  | TATTACAATGAAGAATGCAGAACTGCTCCACCCCGGACCA-----                 | 3104         |
| Db  | 562   | SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer  | 581          |
| QY  | 3105  | GACATCAAGCTTTATCCATTCATATACCATGCAGTTGAGTCTGTGACACAGACCACCAAC  | 3164         |
| Db  | 582   | LysLysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAsp      | 598          |
| QY  | 3165  | CAGCCGGGACAAAGACGGGGGCC                                       | 3188         |
| Db  | 599   | ValProGlyGluGlyAlaSerAla                                      | 606          |
| RESULT 10                                 |   |   |              |
| ID  | Q960N8  | PRELIMINARY;  | PRT; 723 AA. |
| AC  | Q960N8;   |   |              |
| DT  | 01-DEC-2001   | (TrEMBLrel. 19, Created)                                      |              |
| DT  | 01-DEC-2001   | (TrEMBLrel. 19, Last sequence update)                         |              |
| DT  | 01-JUN-2003   | (TrEMBLrel. 24, Last annotation update)                       |              |
| DE  | LD42380p.   |   |              |
| GN  | PARG OR EG:114E2.1 OR CG2864.                                     |   |              |
| OS  | Drosophila melanogaster (Fruit fly).                              |   |              |
| OC  | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;     |   |              |
| OC  | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;        |   |              |
| OC  | Ephydroidea; Drosophilidae; Diptera; Drosophila.                  |   |              |
| OX  | NCBI_TaxID=7227;  |   |              |
| RN  | [1]   |   |              |
| RP  | SEQUENCE FROM N.A.  |   |              |
| RC  | STRAIN=Berkeley;  |   |              |
| RA  | Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,     |   |              |
| RA  | Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R., |   |              |
| RA  | Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J., |   |              |
| RA  | Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K., |   |              |
| RA  | Yu C., Lewis S.E., Rubin G.M., Celniker S.;                       |   |              |
| RL  | Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.           |   |              |
| DR  | EMBL; AY051955; AAK93379.1; -                                     |   |              |
| DR  | FlyBase; FBgn0023216; PARG.                                       |   |              |
| DR  | InterPro; IPR007724; PARG.  |   |              |
| DR  | Pfam; PF05028; PARG; 1.   |   |              |
| SQ  | SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;                |   |              |
| Alignment Scores:                         |   |   |              |
| Pred. No.:                                | 1.65e-73  | Length:   | 723          |
| Score:                                    | 1061.50   | Matches:  | 238          |
| Percent Similarity:                       | 58.66%  | Conservatives:  | 87           |
| Best Local Similarity:                    | 42.96%  | Mismatches:   | 188          |
| Query Match:                              | 14.66%  | Indels:   | 41           |
| DB:                                       | 5   | Gaps:   | 13           |
| US-09-302-812-1 (1-4070) x Q960N8 (1-723) |   |   |              |
| QY  | 1602  | TCTCCAGATAAGAAATGGCTTGGAACCTCTATTGAGGAGATG-----AGGAGATGCCA    | 1655         |
| Db  | 24  | SerLeuAspAspSerTrpArgGlyValSerMetGluAlaIleHisArgAsnArgGlnPro  | 43           |
| QY  | 1656  | AGGTGTGGGATCCGGCTGCTCCCTTGAGACCATTCTGCCAATCACACAGTACTATTCGG   | 1715         |
| Db  | 44  | PheGluLeuGluAsnLeuProProValThrAlaGlyAsnLeuHisArgValMetTyrGln  | 63           |
| QY  | 1716  | GTAGATCTTTTGGGAATAGGAGAAGTTCCTAAACCTTTCCCAACACATTTTAAAGATTG   | 1775         |
| Db  | 64  | -----LeuProIleArgGluThrPro---ProArgProTyrLysSerProGlyLys      | 79           |
| QY  | 1776  | TGGGACAAACAGCATGTTAAGATGCCTTGTTCAGAACAAAACCTGTACCCCTGTGGAAGAT | 1835         |
| Db  | 80  | TrpAspSerGluHisValArgLeuProCysAlaProGluSerLysTyrProArgGluAsn  | 99           |
| QY  | 1836  | GAGAAATGGTGAGCGAGCTGCAGGCGCCGGTGGGAACCTCATTCAGACTGCATCTCTCAAC | 1895         |

|    |      |   |      |
|----|------|---|------|
| Db | 100  | ProAspGlySerThrThrIleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGln  | 119  |
| QY | 1896 | AGGCTCACTCGGCCCCAGAACTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCT   | 1955 |
| Db | 120  | ProIleLysThrCysGluGluLeuGlnAlaAlaIleSerTyrAsnThrThrTyrArg     | 139  |
| QY | 1956 | AAGAAATGGGACTTTACAGCTTTTGATTGTTCTGGGATAAGGTACTAGAAGAACAGAA    | 2015 |
| Db | 140  | AspGlnTrpHisPheArgAlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGlu  | 159  |
| QY | 2016 | GCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAATAATGCACCTCTGCTGCCA  | 2075 |
| Db | 160  | ThrArgValPhePheGluAspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuPro  | 179  |
| QY | 2076 | AATATTGTACCCAGCAATACCACTCCTGAAACAGAGATGAATCATTCATCCACAATG     | 2135 |
| Db | 180  | AspLeuIleGlnSerProValProLeuLeuLysHisLysAsnAlaSerLeuSerLeu     | 199  |
| QY | 2136 | TCACAGGAACAGATTGCCAGTCTTTTAGCTAATAGCTTTCTTCCTGCACGTTTCCACGACG | 2195 |
| Db | 200  | SerGlnGlnGlnIleSerCysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArg  | 219  |
| QY | 2196 | AATGCC---AAGATGAATCAGAGTATTCAGTTATCCAGATATTAATCAATCGGTTG      | 2252 |
| Db | 220  | AsnThrLeuLysArgLysSerGluTyrSerThrPheProAspIleAsnPheAsnArgLeu  | 239  |
| QY | 2253 | TTTGAAGGACGTTTCATCAAGGAACACAGAGAAGCTTAAACGCTCTTCTGCTACTTTAGA  | 2312 |
| Db | 240  | TyrGlnSerThrGlyProAlaValLeuGluLysLeuLysCysIleMetHisTyrPheArg  | 259  |
| QY | 2313 | AGAGTC-----ACAGAGAAAAA-----CCCACTGGGTGGTGACATTCA              | 2354 |
| Db | 260  | ArgValCysProThrGluArgAspAlaSerAsnValProThrGlyValValThrPheVal  | 279  |
| QY | 2355 | AGACAGAGT-----CTTGAAGATTTTCCAGAGTGGGAAGATGTGAAAAAACTCTG---    | 2405 |
| Db | 280  | ArgArgSerGlyLeuProGluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGly  | 299  |
| QY | 2406 | ---ACTCGACTGCATGCTACCTACGAAGTACCATAGAAGAAACGGCCAGGCATGCTA     | 2462 |
| Db | 300  | AspValProLeuHisValAspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeu  | 319  |
| QY | 2463 | CAGGTGGAATTTGCAAAACCGTTTCGTTGGAGGTGGTGTAAACAGTCAGGACTTGTGCA   | 2522 |
| Db | 320  | GlnValAspPheAlaAsnLysTyrLeuGlyGlyValLeuGlyHisGlyCysValGln     | 339  |
| QY | 2523 | GAAGAAATCCGCTTTTAATCAACCTCGAGTTGATTTTACGGCTCTTCACTGAGGTG      | 2582 |
| Db | 340  | GluGluIleArgPheValIleCysProGluLeuLeuValGlyLysLeuPheThrGluCys  | 359  |
| QY | 2583 | CTGGATCACAATGAATGCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGC   | 2642 |
| Db | 360  | LeuArgProPheGluAlaLeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGly  | 379  |
| QY | 2643 | TATGCCGAAACATACCGCTGGGCCCGGAGCCATGAAGACAGGAGCGGAAAGGACGACTGG  | 2702 |
| Db | 380  | TyrAlaGlySerPheGluTrpSerGlyAsnPheGluAspSerThrProArgAspSerSer  | 399  |
| QY | 2703 | CAGAGCGCACGACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGAC   | 2762 |
| Db | 400  | GlyArgArgGlnThrAlaIleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHis  | 419  |
| QY | 2763 | CAGTTTGTGCCCGAGAAGATCAGCGGGAGCTTAACAGGCTTACTGTGGATTCTTCGT     | 2822 |
| Db | 420  | GlnTyrArgGluAspLeuMetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHis  | 439  |
| QY | 2823 | -----CCTGGAGTTTCTTCAGAGAACTGTCTGCAGTGGCTACAGGA                | 2864 |
| Db | 440  | TrpMetValThrProProGly-----ValAlaThrGly                        | 451  |
| QY | 2865 | AACTGGGCTGTGCTGCTTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTG    | 2924 |



Db 452 AsnTrpGlyCysGlyAlaPheGlyAlaPheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMet 471

QY 2925 GCAGCTGCTGTAGCTGAGCGAGACGTGGTTTATTTTCACCTTTTGGGACTCAGAACTGATG 2984

Db 472 ValCysAlaGlnLeuGlyArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArg 491

QY 2985 AGAGACATTTACAGCATGCATACATTCCTCACTGAGAGGAACTGACTGTTGGAGAAAGTA 3044

Db 492 AspAspPheHisGluMetTrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeu 511

QY 3045 TATAAGCTGCTACGATATTACATGAAGAATGCAGAAACTGTCTCCACCCCGGACCA 3104

Db 512 TrpSer---IleLeuArgSerTyrSerArgLeuIleLysGluLysSerSerLysGluPro 530

QY 3105 -----GACATCAAGCTTTATCCATTCCATATACCATCCAGTTGAGTCC 3146

Db 531 ArgGluAsnLysAlaSerLysLysLysLeuTyrAspPheIle-----LysGluGlu 547

QY 3147 TGTACACAGACCACCAACCAGCGCGGACAAAGGACGGGGCC 3188

Db 548 LeuLysLysValArgAspValProGlyGluGlyAlaSerAla 561

RESULT 11

Q9SKB3

ID Q9SKB3 PRELIMINARY; PRT; 548 AA.

AC Q9SKB3; Q94ET7;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative poly (ADP-ribose) glycohydrolase.

GN AT2G31870 OR TEJ.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,

RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,

RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,

RA Fraser C.M., Venter J.C.;

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Panda S., Poirier G.G., Kay S.A.;

RT "TEJ defines a role for poly-ADP-ribosylation in establishing period

RT length of the Arabidopsis circadian oscillator.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC006533; AAD32285.2; -.

DR EMBL; AF394690; AAK72256.1; -.

DR PIR; B84726; B84726.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

KW Hydrolase.

SQ SEQUENCE 548 AA; 62169 MW; 1A79FDA157C3329 CRC64;

Alignment Scores:

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 1.57e-36 | Length:       | 548 |
| Score:                 | 585.00   | Matches:      | 146 |
| Percent Similarity:    | 46.78%   | Conservative: | 65  |
| Best Local Similarity: | 32.37%   | Mismatches:   | 136 |
| Query Match:           | 8.08%    | Indels:       | 104 |
| DB:                    | 10       | Gaps:         | 10  |

US-09-302-812-1 (1-4070) x Q9SKB3 (1-548)

QY 1989 TGGGATAAGGTACTAGAAGAAGCAGCAAGCTCAACACTTGATCAGTCCATCTTGCCCTGAT 2048

Db 79 PheAspGluLeuIleAspGluLysGluSerLysArgTrpPheAspGluIleProAla 98

QY 2049 ATGGTGAAAAATTGCACCTCTGTCTGCCAAATATT----- 2081

Db 99 LeuAlaSerLeuLeuLeuGlnPheProSerLeuLeuGluValHisPheGlnAsnAlaAsp 118

QY 2082 -----TGTACCCAGCCAATACCACCTCCTGAAACAGAAAGATGAATCATTC 2126

Db 119 AsnIleValSerGlyIleLysThrGlyLeuArgLeuLeuAsnSerGlnGlnAlaGlyIle 138

QY 2127 ATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACGTTT 2186

Db 139 ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158

QY 2187 CCACGACGCAATGCCAAGATGAATCAGAGTATTCAGTTATCCAGATATTAACTTCAAT 2246

Db 159 ProAspAspAsnArgGlyAlaLys-----HisLeuProValIleAsnPheAsp 174

QY 2247 CGGTTGTTTGAA-----GGACGTTTCATCAAGGAAACCCAGAGAAGCTTAAAAACG 2294

Db 175 HisLeuPheAlaSerLeuTyrIleSerTyrSerGlnSerGlnGlnSerLysIleArgCys 194

QY 2295 CTCCTTCTGCTACTTTAGAAAGATCAGAGAAAAAACCCTCTGGGTGGTGACATTCCACA 2354

Db 195 IleMetHisTyrPheGluArgPheCysSerCysValProIleGlyIleValSerPheGlu 214

QY 2355 AGACAG-----AGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAAATCCTGACT 2408

Db 215 ArgLysIleThrAlaAlaProAspAlaAspPheTrpSerLysSerAspValSerLeuCys 234

QY 2409 CGACTGCATGTCACTTACGAAGGTACATAGAAGGAAACGGCCAGGCGCATGCTACAGGTG 2468

Db 235 AlaPheLysValHisSerPheGlyLeuIleGluAspGlnProAspAsnAlaLeuGluVal 254

QY 2469 GATTTTGCAAAACCGTTTCGTGGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAGAA 2528

Db 255 AspPheAlaAsnLysTyrLeuGlyGlySerLeuSerArgGlyCysValGlnGluGlu 274

QY 2529 ATCCGCTTTTAAATCAACCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGAT 2588

Db 275 IleArgPheMetIleAsnProGluLeuIleAlaGlyMetLeuPheLeuProArgMetAsp 294

QY 2589 CACAATGAATGCTTATCATCACAGGTACTGACAGTACAGTGAATACACAGGCTATGCC 2648

Db 295 AspAsnGluAlaIleGluIleValGlyAlaGluArgPheSerCysTyrThrGlyTyrAla 314

QY 2649 GAAACATACCGCTGGCGCCGAGCCATGAAGACAGGAGCGAAAGGACGACTGGCAGAGG 2708

Db 315 SerSerPheArgPheAlaGlyGluTyrIleAspLysLysAlaMetAspPropheLysArg 334

QY 2709 CGCAGCACTGAGATCGTCCCATCGACGCCCTC-----CACTTCAGA 2750

Db 335 ArgArgThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLys 354

QY 2751 CGCTACCTCGACCACTTTGTGCCGAGAAGATCAGACGGGAGCTTAAACAAGGCTTACTGT 2810

Db 355 AspIleCys-----LeuLeuArgGluIleAsnLysAlaLeuCys 367

QY 2811 GGATTTCTTCGT----- 2822

Db 368 GlyPheLeuAsnCysSerLysAlaTrpGluHisGlnAsnIlePheMetAspGluGlyAsp 387

QY 2822 ----- 2822

Db 388 AsnGluIleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrThr 407

QY 2822 ----- 2822

Db 408 AlaSerHisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnAsn 427

QY 2823 -----CCTGGAGTTTCTTCAGAGAACCTG-----TCTGCAGTG 2855

Db 428 LeuIleArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspAspGlyVal 447

QY 2856 GCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCTAGACTAAAGCCTTAATA 2915

|   |   |  |              |
|---|---|--|--------------|
| Db  | 448   | AlaThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIle   | 467          |
| QY  | 2916  | CAGATCCGCGAGCTGCTAGCTGAGCGGACGCTGGTT---TATTACACCTTTGGGGAC      | 2972         |
| Db  | 468   | GlnTrpLeuAlaAlaSerGlnThrArgArgProPheIleSerTyrtTrpPheGly---     | 486          |
| QY  | 2973  | TCAGAACTGATGAGAGACATTACAGCATGCATACATTCCTCACTGAGAGGAAACTGACT    | 3032         |
| Db  | 487   | ValGluAlaLeuArgAsnLeuAspGlnValThrLysTrpIleLeuSerHisLysTrpThr   | 506          |
| QY  | 3033  | GTTGGAGAGTATATAAGCTGCTGCTACGATAT                               | 3065         |
| Db  | 507   | ValGlyAspLeuTrpAsnMetMetLeuGluTyr                              | 517          |
| RESULT 12                                 |   |  |              |
| Q8VYAL                                    |   |  |              |
| ID  | Q8VYAL  | PRELIMINARY;   | PRT; 522 AA. |
| AC  | Q8VYAL;   |  |              |
| DT  | 01-MAR-2002 (TrEMBLrel. 20, Created)                                  |  |              |
| DT  | 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)                     |  |              |
| DT  | 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)                   |  |              |
| DE  | Putative poly(ADP-ribose) glycohydrolase.                             |  |              |
| GN  | AT2G31870.  |  |              |
| OS  | Arabidopsis thaliana (Mouse-ear cress).                               |  |              |
| OC  | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;    |  |              |
| OC  | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  |  |              |
| OC  | eurosid II; Brassicales; Brassicaceae; Arabidopsi.                    |  |              |
| OX  | NCBI_TaxID=3702;  |  |              |
| RN  | [1]   |  |              |
| RP  | SEQUENCE FROM N.A.  |  |              |
| RA  | Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,       |  |              |
| RA  | Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,        |  |              |
| RA  | Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J., |  |              |
| RA  | Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,          |  |              |
| RA  | Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,     |  |              |
| RA  | Ecker J., Theologis A., Davis R.W.;                                   |  |              |
| RL  | Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.               |  |              |
| DR  | EMBL; AY072330; AAL61937.1; -.  |  |              |
| DR  | GO; GO:0016787; F:hydrolase activity; IEA.                            |  |              |
| DR  | InterPro; IPR007724; PARG.  |  |              |
| DR  | Pfam; PF05028; PARG; 1.   |  |              |
| KW  | Hydrolase.  |  |              |
| SQ  | SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;                    |  |              |
| Alignment Scores:                         |   |  |              |
| Pred. No.:                                | 1.7e-32   | Length:  | 522          |
| Score:                                    | 533.00  | Matches:   | 132          |
| Percent Similarity:                       | 49.65%  | Conservative:  | 83           |
| Best Local Similarity:                    | 30.48%  | Mismatches:  | 138          |
| Query Match:                              | 7.36%   | Indels:  | 80           |
| DB:                                       | 10  | Gaps:  | 11           |
| US-09-302-812-1 (1-4070) x Q8VYAL (1-522) |   |  |              |
| QY  | 1986  | TTCTGGGATAAGGTACTAGAAGACGAGCTCAACTGTGTATCAGTCCATCTTGCCT        | 2045         |
| Db  | 80  | PhePheAspLysLysIleSerArgGluGluSerAlaAsnPhePheGlyGluValPro      | 99           |
| QY  | 2046  | GATATGGTGAATAATTGCACCTCTGTCTGCCAATAATTTGTACCCAGCCAATACCACTCCTG | 2105         |
| Db  | 100   | AlaLeuCysArgLeuLeuLeuGlnLeuProSerMetLeuGluLysHisTyr-----       | 116          |
| QY  | 2106  | AAACAGAAGATGAATCAT-----  | 2123         |
| Db  | 117   | ---GlnLysAlaAspHisValLeuAspGlyValLysSerGlyLeuArgLeuGlyPro      | 135          |
| QY  | 2124  | -----TCCATCAATGTTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCT           | 2171         |
| Db  | 136   | GlnGluAlaGlyIleValLeuLeuSerGlnGluLeuIleAlaAlaLeuAlaCysSer      | 155          |
| QY  | 2172  | TTCTTCTGCACGTTTCCACGACGCAATGCCAAGATGAATCAGAGTATTCAGTTATCCA     | 2231         |

|           |        |   |              |
|-----------|--------|---|--------------|
| Db        | 156    | PhePheCysLeuPheProGluValAspArgSerLeuLys-----AsnLeuGln         | 171          |
| QY        | 2232   | GATATTAACTTCAATCGGTTGTTTGAA-----GGACGTTTCATCAAGGAAACCA        | 2279         |
| Db        | 172    | GlyIleAsnPheSerGlyLeuPheSerPheProTyrMetArgHisCysThrLysGlnGlu  | 191          |
| QY        | 2280   | GAGAAAGCTTAAAAACGCTCTCTCTACTTCTAGAGAGTACAGAGAAAAAACCCACTGGG   | 2339         |
| Db        | 192    | AsnLysIleLysCysLeuIleHisTyrPheGlyArgIleCysArgTrpMetProThrGly  | 211          |
| QY        | 2340   | TTGGTGACATTCAAGACACAGAGTCTTGAA-----GATTTT                     | 2375         |
| Db        | 212    | PheValSerPheGluArgLysIleLeuProLeuGluTyrHisProHisPheValSerTyr  | 231          |
| QY        | 2376   | CCAGAG-----TGGGAAAGATGTGAAAAAACTCCTGACTCGACTGCATGTCACTTAC     | 2426         |
| Db        | 232    | ProLysAlaAspSerTrpAlaAsnSerValThrProLeuCysSerIleGluIleHisThr  | 251          |
| QY        | 2427   | GAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTCGCAAAACCGTTTC | 2486         |
| Db        | 252    | SerGlyAlaIleGluAspGlnProCysGluAlaLeuGluValAspPheAlaAspGluTyr  | 271          |
| QY        | 2487   | GTTGGAGGTGGTAAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTAAATCAAC   | 2546         |
| Db        | 272    | PheGlyGlyLeuThrLeuSerTyrAspThrLeuGlnGluIleArgPheValIleAsn     | 291          |
| QY        | 2547   | CCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTTATC  | 2606         |
| Db        | 292    | ProGluLeuIleAlaGlyMetIlePheLeuProArgMetAspAlaAsnGluAlaIleGlu  | 311          |
| QY        | 2607   | ATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGCTGGCC   | 2666         |
| Db        | 312    | IleValGlyValGluArgPheSerGlyTyrThrGlyTyrGlyProSerPheGlnTyrAla  | 331          |
| QY        | 2667   | CGGAGCCATGAACACAGGAGCGAAAGGGACGACTGGCAGAGCGCACGACTGAGATCGTC   | 2726         |
| Db        | 332    | GlyAspTyrThrAspAsnLysAspLeuAspIlePheArgArgArgLysThrArgValIle  | 351          |
| QY        | 2727   | GCCATCGACGCCCTCCACTTCAGACGCTACTCTGACGAGTGTGTCGCCGAGAGATCAGA   | 2786         |
| Db        | 352    | AlaIleAspAlaMetPro---AspProGlyMetGlyGlnTyrLysLeuAspAlaLeuIle  | 370          |
| QY        | 2787   | CGGAGCTTAACAAGGCTTACTGTGGATTCTTCGT-----                       | 2822         |
| Db        | 371    | ArgGluValAsnLysAlaPheSerGlyTyrMetHisGlnCysLysTyrAsnIleAspVal  | 390          |
| QY        | 2823   | -----CCTGGAGTTTCTTCA-----                                     | 2837         |
| Db        | 391    | LysHisAspProGluAlaSerSerSerHisValProLeuThrSerAspSerAlaSerGln  | 410          |
| QY        | 2838   | -----GAGAACCTGTCTGCAGTGGCT                                    | 2858         |
| Db        | 411    | ValIleGluSerSerHisArgTrpCysIleAspHisGluGluLysLysIleGlyValAla  | 430          |
| QY        | 2859   | ACAGGAAACTGGGCTGTGTGCTTTTGGGGGTGATGCTAGACTAAAGCCCTTAATACAG    | 2918         |
| Db        | 431    | ThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysIleMetLeuGln  | 450          |
| QY        | 2919   | ATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTT---TATTTCACCTTTGGGACTCA   | 2975         |
| Db        | 451    | TrpLeuAlaIleSerGlnSerGlyArgProPheMetSerTyrTyrThrPheGly---Leu  | 469          |
| QY        | 2976   | GAACTGATGAGAGACATTACAGCATGCATACATTCCTCAGAGAGGAAGAACTGACTGTT   | 3035         |
| Db        | 470    | GlnAlaLeuGlnAsnLeuAsnGlnValIleGluMetValAlaLeuGlnGluMetThrVal  | 489          |
| QY        | 3036   | GGAGAAGTATATAAGCTGCTGCTACGATATTTACAAATGAA                     | 3074         |
| Db        | 490    | GlyAspLeuTrpLysLeuValGluTyrSerSerGlu                          | 502          |
| RESULT 13 |        |   |              |
| Q867X0    |        |   |              |
| ID        | Q867X0 | PRELIMINARY;  | PRT; 781 AA. |

AC Q867X0;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3  
 GN protein) (Corresponding sequence F20C5.1a).  
 GN F20C5.1 OR PME-3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gagnon S.N., Hardy I., Desnoyers S.;  
 RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode  
 RT Caenorhabditis elegans.";  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Matthews P.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY185493; AAO26316.1; -.  
 DR EMBL; Z68161; CAD89735.1; -.  
 DR WormPep; F20C5.1a; CE33775.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 SQ SEQUENCE 781 AA; 89273 MW; 30447D06605CED0A CRC64;

Alignment Scores:

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 1.95e-28 | Length:       | 781 |
| Score:                 | 481.50   | Matches:      | 210 |
| Percent Similarity:    | 38.53%   | Conservative: | 126 |
| Best Local Similarity: | 24.08%   | Mismatches:   | 297 |
| Query Match:           | 6.65%    | Indels:       | 239 |
| DB:                    | 5        | Gaps:         | 37  |

US-09-302-812-1 (1-4070) x Q867X0 (1-781)

|    |      |  |      |
|----|------|--|------|
| QY | 852  | GACACAAATAGTGAGGAGATAGAGATAATCAGCAGTCTTTTGACACATGTAAGCTTGG   | 911  |
| Db | 10   | AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla | 28   |
| QY | 912  | AATGCAAAGCAGACGATGGAA-----GATGAACAGGGCAGAGAGCCAGAGC          | 959  |
| Db | 29   | HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThrGluSer | 48   |
| QY | 960  | CACCAGAAAGTGTGGCAAGGCTTCCCATCCTGCAGAGCCCTGTGCAGGTTGCAGCAGGAG | 1019 |
| Db | 49   | LysGlu-----AspProGluGlu-----ProLys                           | 56   |
| QY | 1020 | GAGACAGACGTGGTGTCCGAGAGCCCTTGTCCGACACTGGCTCTGAGGAT-----      | 1070 |
| Db | 57   | SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu | 76   |
| QY | 1071 | -----GTTGGTACTGGACTGAAATATGCC-----                           | 1094 |
| Db | 77   | AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys | 96   |
| QY | 1095 | -----AACAGATTGAATAGACAAAGAGTAGCTAGGAAAT                      | 1130 |
| Db | 97   | IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn | 116  |
| QY | 1131 | -----TCTCCTCCATTGTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGAT    | 1181 |
| Db | 117  | HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspVal--- | 135  |

|    |      |   |      |
|----|------|---|------|
| QY | 1182 | AATTCCAAAATAAGTTGTGAGGATTCAGAAGCA-----GATGAAGAGACA            | 1226 |
| Db | 136  | ---AlaIleAsnSerAspGluAspGluValLeuGluGluAsnLysGluMet           | 154  |
| QY | 1227 | AGTCCAGGTTTGATGAACAGGAAGAT-----                               | 1253 |
| Db | 155  | ArgAspGlyGluGlnValGlnGlnAspLeuPheAlaAspGlnGluLeuIleGluTyr     | 174  |
| QY | 1254 | -----AGCAGTTCTGCTCAACA  | 1271 |
| Db | 175  | ProGlyIleMetLysAspThrThrThrGlnLeuAspIleThrAspSerGluValGluThr  | 194  |
| QY | 1272 | GCAATATAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAGCGGTCC   | 1331 |
| Db | 195  | AlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer-----               | 208  |
| QY | 1332 | TCTGCTAAGGAGGTGAGATTCGATTACATTTCCAATTGAAGGAGGAGAGTCCGAGCT     | 1391 |
| Db | 209  | -----ThrPheValGlyGluAspSerLys---                              | 216  |
| QY | 1392 | GGAATGAATGATGTAATGCCAAACGACCTGGAAGTACT-----                   | 1430 |
| Db | 217  | -----AsnGlnArgGlnSerGlyThrThrSerAspGluValAspAlaAsp            | 231  |
| QY | 1431 | TCTAGCCTGAATGTAGAGTGCAGAAATTCCTAAGCAACATGGGAGAAAGGATTCCTAAATC | 1490 |
| Db | 232  | SerGlnIleAsnLeuAlaThrLysThrValArg-----ThrSerSer               | 245  |
| QY | 1491 | ACAGATCATTTTCATGAGAGTCCCCAAAGCAGAGGACAAAGAAACAATGTGAAATG      | 1550 |
| Db | 246  | SerSerPheLeuSerThrValSerThrCysGluAlaProAlaLysGlyArgAlaArgMet  | 265  |
| QY | 1551 | AAACATCAAAGAACAGAAAGGAAGATCCCTAATAATCATTCACCTCACCTTTCT        | 1604 |
| Db | 266  | TyrGlnLysGluGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGln     | 285  |
| QY | 1605 | CCAGATAAGAAATGGCTTGGAACTCCTATTGAGGAGATGAGGAGATGCCAAGGTGTGGG   | 1664 |
| Db | 286  | ProAsp-----   | 287  |
| QY | 1665 | ATCCGGCTGCCCTCCCTTGAGACCATCTGCCAATCACACAGTACTATTTCGGGTAGATCTT | 1724 |
| Db | 288  | -----LeuAsnLysValAspProAspArgAsnTyr-----                      | 297  |
| QY | 1725 | TGCGAATAGGAGAGTTCTCTAAACCTTTCCCA---ACACATTTTAAAGATTTTGGGAC    | 1781 |
| Db | 298  | ---ArgTyrCysThrIleProAsn---PheProAlaSerGlnGlyLysLeuArgGluAsp  | 315  |
| QY | 1782 | AACAAGCAT-----GTTAAGATGCCCTTGTTCAGAACAAACTTGTACCTGTGGAA       | 1832 |
| Db | 316  | AsnArgTyrGlyProLysIleValLeuProGlnArgTyrArgGluPhe-----         | 331  |
| QY | 1833 | GATGAGAATGGTGAGCGAGCTGCAGGCGAGCCGGTGGGAACTCATTCAGACTGCCTTCTC  | 1892 |
| Db | 332  | AspSerArgGlyArgArg-----                                       | 337  |
| QY | 1893 | AACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAGTACAAT-----      | 1943 |
| Db | 338  | -----ArgAspSerTyrPheTyrPheLysArgLysLeu                        | 348  |
| QY | 1944 | GTGGCATATTCTAAGAAATGGGACTTTACAGCT-----TTGATT                  | 1982 |
| Db | 349  | AspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHis  | 368  |
| QY | 1983 | GATTCTGGGAT-----AAGGTACTAGAAGAGCAGAAAGCTCAACACTTGTATCAG       | 2033 |
| Db | 369  | AsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyr  | 388  |
| QY | 2034 | TCCATCTTGCCTGATATGGTG-----                                    | 2054 |
| Db | 389  | LysGluMetSerGluLeuValGlyArgGluGluValLeuGluLysPheAlaArgValAla  | 408  |
| QY | 2055 | AAAATGCACTCTGTCTGCCAAATATTGTACCCAGGCCAATACCCTCTCTGAAAACAGAG   | 2114 |



Db 409 ArgIleAlaLysThrAlaGluAspIleLeuProGluArgIleTyrArgLeuValGlyAsp 428  
QY 2115 ATGAATCATTCATCAATGTCACAGGAACAGATGCCAGTCTTTTAGCTAATGCTTTC 2174  
Db 429 Val---GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhe 447  
QY 2175 TTCTGCACGTTTCCACGACGCAATGCCAAGATGAATCAGAGTATTCAGTTATCCAGAT 2234  
Db 448 Phe-----AlaArgProAspSerProPheSer----- 456  
QY 2235 ATTAACCTCAATCGTTGTTTGAAGGACGTTTCATCAAGGAAACCCAGAGAAGCTTAAACG 2294  
Db 457 -----PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeuLysPhe 474  
QY 2295 CTCTTCTGCTACTTTAGAGAGTACACAGAGAAAAAACCCACTGGTGTGTCATTC--- 2351  
Db 475 LeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSerPheArg 494  
QY 2352 ---ACAAGACAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACT 2408  
Db 495 LeuThrLysMetAspLysAspThrPheAsnGluGluTrpLysAspLysLysLeuArgSer 514  
QY 2409 CGACTGCATGTCATTAC-----GAAGTACCATAGAAGAAACGGCCAGGCATGCTA 2462  
Db 515 LeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAlaLeu---CysThr 533  
QY 2463 CAGGTGGATTTTGAAACCGTTTCGTTGGAGTGGTGTAAACAGTGCAGGACTTGTGCAA 2522  
Db 534 GlnValAspPheAlaAsnGluHisLeuGlyGlyValLeuAsnHisGlySerValGln 553  
QY 2523 GAAGAAATCCGCTTTTAAATCAACCTGAGTTGATGTTTTCACGGCTCTTCACTGAGGTG 2582  
Db 554 GluGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeuCysGluLys 573  
QY 2583 CTGGATCAACAATGATGCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGC 2642  
Db 574 MetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSerTyrThrGly 593  
QY 2643 TATGCCGAAACATACCGCTGGGCC-----CGGAGCCATGAAGACAGGAGC 2687  
Db 594 TyrGlyHisThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsn 613  
QY 2688 GAA---AGGACAGCTGGCAGAGGCGCAGCTGAGATCGTCCCATCGCCCTCCAC 2744  
Db 614 GluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeu 633  
QY 2745 TTCAGACGC-----TACCTCGACCATGTTGTGCCCGAGAAAGATCAGACGG 2789  
Db 634 PheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsnIleArg 653  
QY 2790 GAGCTTAACAAGCTTACTGTGATTTCTTCGTCCTGGAGTTCTTTCAGAGAACCTGTCT 2849  
Db 654 GluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThrAsnIle--- 672  
QY 2850 GCAGTGGCTACAGGAACCTGGGCTGTGGTGCCTTTGGGGTGATGCTAGACTAAAGGCC 2909  
Db 673 ProIleValThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPhe 692  
QY 2910 TTAATACAGATCCTGGCAGCTGTAGCTGAGCGAGACGTGGTTTATTTTCACTTTGGG 2969  
Db 693 IleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCysSerPheGly 712  
QY 2970 GACTCAGAACTGATGAGACATTTACAGCATGCATATTCCTCACTGAGAGGAAACTG 3029  
Db 713 GluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGlnLysAspVal 732  
QY 3030 ACTGTTGGAGAGTATATAGCTGCTGCTACGATATTAATGAAGATGCAGAAACTGC 3089  
Db 733 ThrLeuGlyMetLeuPheSerMetIle-----AsnAsnThr 744  
QY 3090 TCCACCCCGGACGACATCAAGCTTTATCCATTC 3125

Db 745 GlyLeuProHisLysHisPheGluPheTyrValPhe 756  
RESULT 14  
Q19637  
ID Q19637 PRELIMINARY; PRT; 764 AA.  
AC Q19637;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose  
metabolism enzyme-3 short form).  
GN F20C5.1 OR PME-3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matthews P.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.  
RX SEQUENCE FROM N.A.  
MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode  
Caenorhabditis elegans.";  
RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; Z68161; CAA92299.2; --.  
DR EMBL; AY185494; AAC26317.1; --.  
DR PIR; T21138; T21138.  
DR WormPep; F20C5.1b; CE32867.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 764 AA; 87441 MW; B391703B1D0CB79A CRC64;  
Alignment Scores:  
Pred. No.: 6.19e-28 Length: 764  
Score: 475.00 Matches: 208  
Percent Similarity: 37.88% Conservative: 120  
Best Local Similarity: 24.02% Mismatches: 294  
Query Match: 6.56% Indels: 244  
DB: 5 Gaps: 36  
US-09-302-812-1 (1-4070) x Q19637 (1-764)  
QY 852 GACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAAGCTTGGC 911  
Db 10 AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla 28  
QY 912 AATGCAAAGCAGACGATGGAA-----GATGAACAGGGCCAGAGAGCCAGAGC 959  
Db 29 HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThrGluSer 48  
QY 960 CACCAGAAAGTGTGGCAAGGCTTGCCATCCTGCAGAAAGCCTGTGCAGGGTGTGCAGCAGGAG 1019  
Db 49 LysGlu-----AspProGluGlu-----ProLys 56  
QY 1020 GAGACAGACGCTGTGTCCGAGAGCCCTTCTTCGGACACTGGCTCTGAGGAT----- 1070  
Db 57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76  
QY 1071 -----GTTGGTACTGGACTGAAAAATGCC----- 1094  
Db 77 AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96  
QY 1095 -----AACAGATTGAATAGACAAAGAGTAGTCTAGGAAAT 1130



Db 97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn 116  
QY 1131 -----TCTCCTCCATTGAGAAAGAAAGTGAACCTGAGTACCAATGGATGTA--- 1178  
Db 117 HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspValAla 136  
QY 1179 -----GATAATTCCAAAAATAGTTGTCAGGAT 1205  
Db 137 IleAsnSerAspGluAspAspGluLeuValLeuGluAsnAsnLysGluMetArgAsp 156  
QY 1206 TCAGAAAGCAGATGAAGACAAAGTCCAGGTTT-----GATGAACAGGAA----- 1250  
Db 157 GlyGluGlnValGlnGlnLeuSerGlnAspLeuPheAlaAspAspGlnGluLeuIleGlu 176  
QY 1251 -----GATAGCAGTTCTGCTCAA 1268  
Db 177 TyrProGlyIleMetLysAspThrThrThrGlnLeuAspIleThrAspSerGluValGlu 196  
QY 1269 ACAGCAAATAAACCTTCAAGGTTCCAAACCAAGAGAAAGCTGACACTGAGTTGAGGAAGCGG 1328  
Db 197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer----- 211  
QY 1329 TCCTCTGCTAAGGAGGTGAGATTTCGATTACATTTCCAAATTGAAGGAGGAGAGAGTCTGA 1388  
Db 212 -----ThrPheValGlyGluAspSerLys 219  
QY 1389 GCTGGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAG 1448  
Db 220 Ala-----ThrLysThrValArgThrSerSerSerPheLeuSerThr--- 234  
QY 1449 TGCAGAAATTCTAAGCAACATGGGAGAAAGGATTCTAAATACACAGATCATTTTCATGAGA 1508  
Db 234 ----- 234  
QY 1509 GTGCCCAAAGCAGAGACAAAGAAAGAAACATGTGAATGAAACATCAACAGAAACAGAA 1568  
Db 235 ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254  
QY 1569 AGGAAGATCCCTAAATACATTCACCTCACCTTTCT-----CCAGATAAGAAATGGCTT 1622  
Db 255 LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp----- 270  
QY 1623 GGAACCTCCTATTGAGGAGATGAGGAGAAATGCCAAGGTGGGATCCGGTGCCTCCCTTG 1682  
Db 271 -----LeuAsnLysVal 274  
QY 1683 AGACCATCTGCCAATCACACAGTGACTATTTCGGGTAGACTTTTTCGGAATAGGAGAAGTT 1742  
Db 275 AspProAspArgAsnTyr-----ArgTyrCysThrIle 285  
QY 1743 CCTAAACCTTTCCCA---ACACATTTTAAAGATTGTGGGACAAACAGCAT----- 1790  
Db 286 ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304  
QY 1791 GTTAAGATGCCTTGTTCAGAAACAAACTTGTACCCCTGTGGAAGATGAGAAATGGTGACCGA 1850  
Db 305 IleValLeuProGlnArgTyrArgGluPhe-----AspSerArgGlyArgArg 320  
QY 1851 GCTGCAGGCAGCCGGTGGGAAGTCACTTCAGACTGCATCTCTCAACAGGCTCACTCGGCC 1910  
Db 320 ----- 320  
QY 1911 CAGAACTGAAGGATGCTATTCTGAAGTACAAT-----GTGGCATATTTCAAGAAA 1961  
Db 321 -----ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337  
QY 1962 TGGGACTTACAGCT-----TTGATTGATTCTTGGGAT----- 1994  
Db 338 TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTyrGluPheAsp 357  
QY 1995 -----AAGGTACTAGAAGACGAGCTCAACACTTGATGATCAGTCCATCTTGCCTGATATG 2051  
Db 358 ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerSerGluLeu 377

QY 2052 GTG-----AAAATTGCACCTCTGTCTG 2072  
Db 378 ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAlaLysThrAla 397  
QY 2073 CCAAAATATTTGTACCCAGCAATATACCACTCCTGAAACAGAAAGATGAATCCATCCATCACA 2132  
Db 398 GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThr 416  
QY 2133 ATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGACAGTTCCTCCACGA 2192  
Db 417 LeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431  
QY 2193 CGCAATGCCAAGATGAAATCAGAGTATTCAGATTATCCAGATATTAATCAATCGGTTG 2252  
Db 432 -----AlaArgProAspSerPropheSer-----PheCysArgIle 443  
QY 2253 TTTGAAGGACGTTTCATCAAGGAAACCCAGAGAGCTTAAACCGCTCTTCTGCTACTTTAGA 2312  
Db 444 LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463  
QY 2313 AGAGTCACAGAGAAAAAACCCACTGGGTGGTGACATTC-----ACAAGACAGAGTCTT 2366  
Db 464 LysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483  
QY 2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAAATCCTGACTCGACTGCTGCTCACTTAC 2426  
Db 484 AspThrPheAsnGluGluTyrLysAspLysLysLeuArgSerLeuProGluValGluPhe 503  
QY 2427 -----GAAGGTACCATAGAGGAAACCGCCAGGCGCATGCTACAGGTGGATTTTGCAAAAC 2480  
Db 504 PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522  
QY 2481 CGTTTCGTTGAGGTGGTGTAAACCAGTGCAGGACTTGTGCAAGAGAAATCCGCTTTTATA 2540  
Db 523 GluHisLeuGlyGlyValLeuAsnHisGlySerValGlnGluGluIleArgPheLeu 542  
QY 2541 ATCAACCTTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600  
Db 543 MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuGluAla 562  
QY 2601 CTTATCATCACAGGTACTGAGCAGTACAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660  
Db 563 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLys 582  
QY 2661 TGGGCC-----CGGAGCCATGAAGACAGGAGCGAA---AGGACGACTGG 2702  
Db 583 TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPhe 602  
QY 2703 CAGAGGCGCACGACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGC----- 2753  
Db 603 GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu 622  
QY 2754 -----TACCTCAGCAGTTTGTGCCCCGAGAGATCAGACGGGAGCTTAACAAGGCTTAC 2807  
Db 623 AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSer 642  
QY 2808 TGTGGATTTCTTCGTCCTGGAGTTTCTTTCAGAGAACCTGTCTGACGTGGCTACAGGAAAC 2867  
Db 643 IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTyr 661  
QY 2868 TGGGGCTGTGGTGCCTTTGGGGGTGATGCTAGACTAAAGCCCTTAATACAGATCCTGGCA 2927  
Db 662 TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681  
QY 2928 GCTGCTGTAGCTGAGCGAGACGTTGTTTATTTCACCTTTGGGACTCAGAACTGATGAGA 2987  
Db 682 AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAla 701  
QY 2988 GACATTTACAGCATCATACATTCCTCCTCAGTGAAGGAAACCTGACTGTTGGAGAGTATAT 3047  
Db 702 LysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyMetLeuPhe 721

|   |  |  |                 |      |  |
|---|--|--|-----------------|------|--|
| QY  | 3048   | AAGCTGCTGCTACGATATTACAATGAAGAATGCAGAAACTGCTCCACCCCGGACCAGAC  | 3107            |      |  |
| Db  | 722  | SerMetile-----AsnAsnThrGlyLeuProHisLysHis                    | 733             |      |  |
| QY  | 3108   | ATCAAGCTTTATCCATTC   | 3125            |      |  |
| Db  | 734  | PheGluPheTyrValPhe   | 739             |      |  |
| RESULT 15                                 |  |  |                 |      |  |
| Q9N5L4                                    | PRELIMINARY; PRT; 485 AA.  |  |                 |      |  |
| ID  | Q9N5L4   |  |                 |      |  |
| AC  | Q9N5L4;  |  |                 |      |  |
| DT  | 01-OCT-2000  | (TrEMBLrel. 15, Created)                                     |                 |      |  |
| DT  | 01-MAR-2003  | (TrEMBLrel. 23, Last sequence update)                        |                 |      |  |
| DT  | 01-JUN-2003  | (TrEMBLrel. 24, Last annotation update)                      |                 |      |  |
| DE  | Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme     |  |                 |      |  |
| DE  | 4).  |  |                 |      |  |
| GN  | H23L24.5 OR PME-4.   |  |                 |      |  |
| OS  | Caenorhabditis elegans.  |  |                 |      |  |
| OC  | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; |  |                 |      |  |
| OC  | Rhabditidae; Peloderinae; Caenorhabditis.                            |  |                 |      |  |
| OX  | NCBI_TaxID=6239;   |  |                 |      |  |
| RN  | [1]  |  |                 |      |  |
| RP  | SEQUENCE FROM N.A.   |  |                 |      |  |
| RC  | STRAIN=Bristol N2;   |  |                 |      |  |
| RX  | MEDLINE=99069613; PubMed=9851916;                                    |  |                 |      |  |
| RA  | Waterston R.;  |  |                 |      |  |
| RT  | "Genome sequence of the nematode C. elegans: a platform for          |  |                 |      |  |
| RT  | investigating biology. The C. elegans Sequencing Consortium."        |  |                 |      |  |
| RL  | Science 282:2012-2018(1998).   |  |                 |      |  |
| RN  | [2]  |  |                 |      |  |
| RP  | SEQUENCE FROM N.A.   |  |                 |      |  |
| RC  | STRAIN=Bristol N2;   |  |                 |      |  |
| RA  | Abu-Threideh J., Lehnert L.;   |  |                 |      |  |
| RT  | "The sequence of C. elegans cosmid H23L24.";                         |  |                 |      |  |
| RL  | Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.             |  |                 |      |  |
| RN  | [3]  |  |                 |      |  |
| RP  | SEQUENCE FROM N.A.   |  |                 |      |  |
| RC  | STRAIN=Bristol N2;   |  |                 |      |  |
| RA  | Waterston R.;  |  |                 |      |  |
| RL  | Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.             |  |                 |      |  |
| RN  | [4]  |  |                 |      |  |
| RP  | SEQUENCE FROM N.A.   |  |                 |      |  |
| RA  | Gagnon S.N., Hardy I., Desnoyers S.;                                 |  |                 |      |  |
| RT  | "Molecular characterization of poly ADP-ribose glycohydrolases in    |  |                 |      |  |
| RT  | Caenorhabditis elegans."   |  |                 |      |  |
| RL  | Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.             |  |                 |      |  |
| DR  | EMBL; AC006662; AAF39896.2; -.                                       |  |                 |      |  |
| DR  | EMBL; AF548468; AAN40699.1; -.                                       |  |                 |      |  |
| DR  | WormPep; H23L24.5; CE32685.  |  |                 |      |  |
| DR  | InterPro; IPR007724; PARG.   |  |                 |      |  |
| DR  | Pfam; PF05028; PARG; 1.  |  |                 |      |  |
| KW  | Hypothetical protein.  |  |                 |      |  |
| SQ  | SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;                   |  |                 |      |  |
| Alignment Scores:                         |  |  |                 |      |  |
| Pred. No.:                                | 9.08e-28   | Length:  | 485             |      |  |
| Score:                                    | 472.00   | Matches:   | 143             |      |  |
| Percent Similarity:                       | 47.18%   | Conservative:  | 83              |      |  |
| Best Local Similarity:                    | 29.85%   | Mismatches:  | 165             |      |  |
| Query Match:                              | 6.52%  | Indels:  | 88              |      |  |
| DB:                                       | 5  | Gaps:  | 17              |      |  |
| US-09-302-812-1 (1-4070) x Q9N5L4 (1-485) |  |  |                 |      |  |
| QY  | 1866   | TGGGAACCTATTTCAGACTGCACTTCTCAACAGGCTCACTCGG----              | CCCCAGAACCTGAAG | 1922 |  |
| Db  | 46   | PheGluLeuLeuGluThrGlyValSerGlnGlnTrpArgAsnCysAspGlnAsnLeuPhe | 65              |      |  |
| QY  | 1923   | GATGCTATTCTGAAG-----TACAAATGTGGCATATTCT-                     | -----           | 1955 |  |
| Db  | 66   | AsnGluTyrLeuLysThrTyrLysAsnGlyGlyTyrSerGlnPheGluAspLeuLeuPhe | 85              |      |  |

|    |      |   |      |  |  |
|----|------|---|------|--|--|
| QY | 1956 | -----AAGAAATGGGACTTTACAGCTTTTGATTGATTTC                         | 1988 |  |  |
| Db | 86   | LysIleTrpGlyTyrSerGluGluLysGluArgPheAspLeuProAlaLeuLysSerPhe    | 105  |  |  |
| QY | 1989 | TGGGATAAGGTACTAGAA-----GAAGCAGAAGCTCAACACTTGTATCAGTCCATC        | 2039 |  |  |
| Db | 106  | TyrArgLysMetSerGluIleValGlyGluAspGlu-----Val                    | 118  |  |  |
| QY | 2040 | TTGCCTGATATGGTGAATAATTGCACTCTGTCTGCCAAATAATTGTACCCAGCCAATACCA   | 2099 |  |  |
| Db | 119  | LeuGluLysLeuAlaAArgLeuValArgIleThrLysSerAlaCys-----Glu          | 134  |  |  |
| QY | 2100 | CTCTGAAACAGAGATGAAT-----CATTCATCACAATGTCACAG                    | 2141 |  |  |
| Db | 135  | ValLeuProGluLysIleTyrArgLeuValGlyAspIleGluSerAlaThrPheSerHis    | 154  |  |  |
| QY | 2142 | GAACAGATTGCCAGTCTTTTAGCTAATGCTTCTTCTGCACGTTTCCACGCAATGCC        | 2201 |  |  |
| Db | 155  | IleGlnCysAlaSerLeuIleAlaTrpMetPheSerAspThrProArg-----           | 171  |  |  |
| QY | 2202 | AAGATGAAATCAGAGTATTCAGTTATCCAGTATCCAGATATTAACCTCAATCGTGTGTAAGGA | 2261 |  |  |
| Db | 172  | -----LeuSerPheIleIleLeuGlnLys                                   | 180  |  |  |
| QY | 2262 | CGTTCAATCAAGGAAACACAGAGAGCTTAAACGCTCTTCTGCTACTCGACTGCATGTC      | 2321 |  |  |
| Db | 181  | ThrThrCysValAlaValGluLysLeuLysPheLeuPheThrTyrPheAspLysMetSer    | 200  |  |  |
| QY | 2322 | GAGAAAAACCCACTGGGTTGGTGACATTACACA-----AGACAGAGT                 | 2363 |  |  |
| Db | 201  | IleAspProProIleGlyAlaValSerPheArgLysMetArgIleThrHisLysGlnTyr    | 220  |  |  |
| QY | 2364 | CTTGAAGATTTTCCAGAGTGGGAA--AGATGTGAAAAAACCCTGACTCGACTGCATGTC     | 2420 |  |  |
| Db | 221  | LeuGluAsn-----TrpLysLeuArgGluThrAsnLeuLeuProAspValGlnVal        | 237  |  |  |
| QY | 2421 | ACTTACGAAGTACCATAGAAGGAACGGCCAGGCGATGCTACAGGTGGATTGTGCAAC       | 2480 |  |  |
| Db | 238  | PheAspLysMetSerIleGluGluThrAlaLeu---CysThrGlnIleAspPheAlaAsn    | 256  |  |  |
| QY | 2481 | CGTTTCGTTGGAGGTGGTGTAAACCAGTGCAGGACTTGTCAAGAGAAATCCGCTTTTA      | 2540 |  |  |
| Db | 257  | LysArgLeuGlyGlyGlyValLeuLysGlyGlyAlaValGlnGluGluIleArgPheMet    | 276  |  |  |
| QY | 2541 | ATCAACCTCAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT    | 2600 |  |  |
| Db | 277  | MetCysProGluMetMetValAlaIleLeuLeuAsnAspValThrGlnAspLeuGluAla    | 296  |  |  |
| QY | 2601 | CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC    | 2660 |  |  |
| Db | 297  | IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrSerAsnThrLeuLys    | 316  |  |  |
| QY | 2661 | TGGGCCCCG-----AGCCATGAAGACAGGAGCGAAGGACGACTGG                   | 2702 |  |  |
| Db | 317  | TrpAlaLysIleThrProLysHisSerAlaGlnAsnAsnSerPheArgAspGlnPhe       | 336  |  |  |
| QY | 2703 | CAGAGGCGCACGACTGAGATCGTCGCCATCGACGCCCTCCAC-----TTC              | 2747 |  |  |
| Db | 337  | GlyArgLeuGlnThrGluThrValAlaIleAspAlaValArgAsnAlaGlyThrProLeu    | 356  |  |  |
| QY | 2748 | AGACGCTACCTCGACCAAGTTTGTGCCGAGAAGATCAGACGGGAGCTTAACAGGCTTAC     | 2807 |  |  |
| Db | 357  | GluCysLeuLeuAsnGlnLeuThrThrGluLysLeuThrArgGluValArgLysAlaAla    | 376  |  |  |
| QY | 2808 | TGTGGATTCTTCGTCCTGGAGTTTCTTCAGAGAACCTGTCTGCGAGTGGCTACAGGAAAC    | 2867 |  |  |
| Db | 377  | IleGlyPheLeuSerAlaGlyAspGlyPheSerLysIle---ProValValSerGlyTrp    | 395  |  |  |
| QY | 2868 | TGGGGCTGTGTGCTTTGGGGGTGATGCTAGACTAAAGCCCTTAATACAGATCCTGGCA      | 2927 |  |  |
| Db | 396  | TrpGlyCysGlyAlaPheArgGlyAsnLysProLeuLysPheLeuIleGlnValIleAla    | 415  |  |  |
| QY | 2928 | GCTGCTGTAGCTGAGCGAGACGCTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGA   | 2987 |  |  |

Thu May 27 09:55:53 2004

Db 416 CysGlyIleSerAspArgProLeuGlnPheCysThrPheGlyAspThrGluLeuAlaLys 435  
QY 2988 GACATTTACAGCATGCATACATTCTCACTGAGAGGAACTGACTGTGTTGGAGAGTATAT 3047  
Db 436 LysCysGluGluMetMetThrLeuPheArgAsnAsnValArgThrGlyGlnLeuPhe 455  
QY 3048 AAGCTGCTGCTACGATATTACAATGAAGATGCAGAACTGCTCCACCCCGGACCAGAC 3107  
Db 456 LeuIleIle-----AsnSerIleGlyProPro 464  
QY 3108 ATCAAGCTT---TATCCATTTCATATACCATGCGAGTTGAGTCCTGTACACAGACCACC 3161  
Db 465 LeuAsnTyrSerGluGlnTyrValPheAspAlaIleArgAlaLysIleAsnSerThr 483

Search completed: May 26, 2004, 17:07:11  
Job time : 282.999 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 31.3065 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-2  
Perfect score: 5184  
Sequence: 1 MSAGPGCEPCTKRPRWDA...YHAVESCTQTTNPGQRTGA 977

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description      |
|------------|--------|-------------|--------|-------|------------------|
| 1          | 5184   | 100.0       | 977    | 4     | US-09-302-812-2  |
| 2          | 5184   | 100.0       | 977    | 4     | US-09-511-477-2  |
| 3          | 5184   | 100.0       | 977    | 4     | US-09-511-507-2  |
| 4          | 4635.5 | 89.4        | 976    | 4     | US-09-302-812-4  |
| 5          | 4635.5 | 89.4        | 976    | 4     | US-09-511-477-4  |
| 6          | 4635.5 | 89.4        | 976    | 4     | US-09-511-507-4  |
| 7          | 4332.5 | 83.6        | 968    | 4     | US-09-302-812-6  |
| 8          | 4332.5 | 83.6        | 968    | 4     | US-09-511-477-6  |
| 9          | 4332.5 | 83.6        | 968    | 4     | US-09-511-507-6  |
| 10         | 1063.5 | 20.5        | 768    | 4     | US-09-302-812-8  |
| 11         | 1063.5 | 20.5        | 768    | 4     | US-09-511-477-8  |
| 12         | 1063.5 | 20.5        | 768    | 4     | US-09-511-507-8  |
| 13         | 465    | 9.0         | 726    | 4     | US-09-302-812-10 |
| 14         | 465    | 9.0         | 726    | 4     | US-09-511-477-10 |
| 15         | 465    | 9.0         | 726    | 4     | US-09-511-507-10 |
| 16         | 167    | 3.2         | 31     | 4     | US-09-302-812-11 |
| 17         | 167    | 3.2         | 31     | 4     | US-09-511-477-11 |
| 18         | 167    | 3.2         | 31     | 4     | US-09-511-507-11 |
| 19         | 155.5  | 3.0         | 1804   | 4     | US-09-362-336A-4 |
| 20         | 153    | 3.0         | 913    | 4     | US-08-971-089-4  |
| 21         | 152    | 2.9         | 914    | 1     | US-08-484-105-2  |
| 22         | 152    | 2.9         | 914    | 1     | US-08-484-106-2  |
| 23         | 149    | 2.9         | 29     | 4     | US-09-302-812-12 |
| 24         | 149    | 2.9         | 29     | 4     | US-09-511-477-12 |
| 25         | 149    | 2.9         | 29     | 4     | US-09-511-507-12 |
| 26         | 147    | 2.8         | 26     | 4     | US-09-302-812-19 |
| 27         | 147    | 2.8         | 26     | 4     | US-09-511-477-19 |

|    |       |     |       |   |                     |                    |
|----|-------|-----|-------|---|---------------------|--------------------|
| 28 | 147   | 2.8 | 26    | 4 | US-09-511-507-19    | Sequence 19, Appl  |
| 29 | 146.5 | 2.8 | 1187  | 1 | US-08-320-559-28    | Sequence 28, Appl  |
| 30 | 146.5 | 2.8 | 1187  | 3 | US-08-545-860D-28   | Sequence 28, Appl  |
| 31 | 146.5 | 2.8 | 1187  | 5 | PCT-US94-04496-28   | Sequence 28, Appl  |
| 32 | 146.5 | 2.8 | 1210  | 1 | US-08-320-559-26    | Sequence 26, Appl  |
| 33 | 146.5 | 2.8 | 1210  | 3 | US-08-545-860D-26   | Sequence 26, Appl  |
| 34 | 146.5 | 2.8 | 1210  | 5 | PCT-US94-04496-26   | Sequence 26, Appl  |
| 35 | 143.5 | 2.8 | 717   | 3 | US-08-910-925-1     | Sequence 1, Appli  |
| 36 | 143.5 | 2.8 | 1177  | 4 | US-09-134-001C-5106 | Sequence 5106, Ap  |
| 37 | 143   | 2.8 | 703   | 3 | US-08-910-925-4     | Sequence 4, Appli  |
| 38 | 143   | 2.8 | 1783  | 4 | US-09-362-336A-2    | Sequence 2, Appli  |
| 39 | 142.5 | 2.7 | 1312  | 4 | US-09-345-882-29    | Sequence 29, Appl  |
| 40 | 141.5 | 2.7 | 1805  | 1 | US-07-853-913-2     | Sequence 2, Appli  |
| 41 | 141   | 2.7 | 778   | 6 | 5198347-4           | Patent No. 5198347 |
| 42 | 140.5 | 2.7 | 951   | 4 | US-09-125-635-8     | Sequence 8, Appli  |
| 43 | 140.5 | 2.7 | 1420  | 4 | US-09-125-635-4     | Sequence 4, Appli  |
| 44 | 139.5 | 2.7 | 10182 | 4 | US-09-134-001C-3159 | Sequence 3159, Ap  |
| 45 | 137.5 | 2.7 | 879   | 4 | US-09-914-259-38    | Sequence 38, Appl  |

ALIGNMENTS

RESULT 1

US-09-302-812-2

; Sequence 2, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZY  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-2

Query Match 100.0%; Score 5184; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |     |  |     |
|----|-----|--|-----|
| QY | 1   | MSAGPGCEPCTKRPRWDA...YHAVESCTQTTNPGQRTGA 60                      | 60  |
| Db | 1   | MSAGPGCEPCTKRPRWDA...YHAVESCTQTTNPGQRTGA 60                      | 60  |
| QY | 61  | GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSENNNTREESMMSSVQKDNFY 120  | 120 |
| Db | 61  | GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSENNNTREESMMSSVQKDNFY 120  | 120 |
| QY | 121 | QHNMEKLENVSQLGFDKSPVEKGTQYLKQHTAAMCKWQNEGPHSERLLESEPPAVTLVP 180  | 180 |
| Db | 121 | QHNMEKLENVSQLGFDKSPVEKGTQYLKQHTAAMCKWQNEGPHSERLLESEPPAVTLVP 180  | 180 |
| QY | 181 | EQFSNANVDQSSPKDDHSDTNSSESDNQQLFTHVKLANAKQTMEDEQGREARSHQKCGK 240  | 240 |
| Db | 181 | EQFSNANVDQSSPKDDHSDTNSSESDNQQLFTHVKLANAKQTMEDEQGREARSHQKCGK 240  | 240 |
| QY | 241 | ACHPAEACAGCQEQEETDVVSESLPLSDTGSSEDTGLKNANRLNRQESSLGNSPPFKESE 300 | 300 |
| Db | 241 | ACHPAEACAGCQEQEETDVVSESLPLSDTGSSEDTGLKNANRLNRQESSLGNSPPFKESE 300 | 300 |
| QY | 301 | PESPMVDVNSKNSCQDSEADEFTSPGFDEQEDSSAQATANKPSRFQPREADTELKRSSA 360  | 360 |



||||| 301 PESPMVDVNSKNSQDSEADEETSPGFDEQEDSSSAQTANKPSRFPQREADTELKRSSA 360  
361 KGGEIRLHFQFEGGESRAGMNDVNAKRPSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
361 KGGEIRLHFQFEGGESRAGMNDVNAKRPSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
421 AEDKRKEQCEMKHQRTERKIPKYPHLSPDKKWLGTPTEEMRMRPCGIRLPPLRPSAN 480  
421 AEDKRKEQCEMKHQRTERKIPKYPHLSPDKKWLGTPTEEMRMRPCGIRLPPLRPSAN 480  
481 HTVTIRVDLLRIGEVKPPFTHPKDLWDNKHVKMPCSEQNLYPVEDENGERRAAGSRWELI 540  
481 HTVTIRVDLLRIGEVKPPFTHPKDLWDNKHVKMPCSEQNLYPVEDENGERRAAGSRWELI 540  
541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEAOHLYQSILPDMVK 600  
541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEAOHLYQSILPDMVK 600  
601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660  
601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660  
661 NFNRLFEGSSSRKPEKLTFLCYFRRVTEKKPTGLVTFTRQSLDPPPEWERCEKLLTRLH 720  
661 NFNRLFEGSSSRKPEKLTFLCYFRRVTEKKPTGLVTFTRQSLDPPPEWERCEKLLTRLH 720  
721 VTYEGTIEGNGQGMLOVDFAANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTVLDHNE 780  
721 VTYEGTIEGNGQGMLOVDFAANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTVLDHNE 780  
781 CLIIITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRTEIIVDAIDALHFRRYLDQFVPE 840  
781 CLIIITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRTEIIVDAIDALHFRRYLDQFVPE 840  
841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
901 FTFGDSELMDIYSMHTFLTERKLTIVGEVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 960  
901 FTFGDSELMDIYSMHTFLTERKLTIVGEVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 960  
961 VESCTQTTNQPGORTGA 977  
961 VESCTQTTNQPGORTGA 977

RESULT 2  
US-09-511-477-2  
; Sequence 2, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-2

Query Match 100.0%; Score 5184; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSAGPGCEPCTKRPRWDAATAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSGCGAL 60  
Db 1 MSAGPGCEPCTKRPRWDAATAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSGCGAL 60  
QY 61 GRAGQHRGSATSLVFKQKITTSWMDTKGKTIVSESLHSKENNNTREESMMSSVQKDNFY 120  
Db 61 GRAGQHRGSATSLVFKQKITTSWMDTKGKTIVSESLHSKENNNTREESMMSSVQKDNFY 120  
QY 121 QHNMEKLENVSQGLGFDKSPVEKGTQYLKQHQTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
Db 121 QHNMEKLENVSQGLGFDKSPVEKGTQYLKQHQTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
QY 181 EQFSNANVDQSSPKDDHSDTNSESRDNQQFLTHVVLANAKQTMEDQGREARSHQKCGK 240  
Db 181 EQFSNANVDQSSPKDDHSDTNSESRDNQQFLTHVVLANAKQTMEDQGREARSHQKCGK 240  
QY 241 ACHPAEACAGCQOEETDVVSESPSLDTGSEDVGTGLKANRLNRQESSLNSPPFEKES 300  
Db 241 ACHPAEACAGCQOEETDVVSESPSLDTGSEDVGTGLKANRLNRQESSLNSPPFEKES 300  
QY 301 PESPMVDVNSKNSQDSEADEETSPGFDEQEDSSSAQTANKPSRFPQREADTELKRSSA 360  
Db 301 PESPMVDVNSKNSQDSEADEETSPGFDEQEDSSSAQTANKPSRFPQREADTELKRSSA 360  
QY 361 KGGEIRLHFQFEGGESRAGMNDVNAKRPSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
Db 361 KGGEIRLHFQFEGGESRAGMNDVNAKRPSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
QY 421 AEDKRKEQCEMKHQRTERKIPKYPHLSPDKKWLGTPTEEMRMRPCGIRLPPLRPSAN 480  
Db 421 AEDKRKEQCEMKHQRTERKIPKYPHLSPDKKWLGTPTEEMRMRPCGIRLPPLRPSAN 480  
QY 481 HTVTIRVDLLRIGEVKPPFTHPKDLWDNKHVKMPCSEQNLYPVEDENGERRAAGSRWELI 540  
Db 481 HTVTIRVDLLRIGEVKPPFTHPKDLWDNKHVKMPCSEQNLYPVEDENGERRAAGSRWELI 540  
QY 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEAOHLYQSILPDMVK 600  
Db 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEAOHLYQSILPDMVK 600  
QY 601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660  
Db 601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660  
QY 661 NFNRLFEGSSSRKPEKLTFLCYFRRVTEKKPTGLVTFTRQSLDPPPEWERCEKLLTRLH 720  
Db 661 NFNRLFEGSSSRKPEKLTFLCYFRRVTEKKPTGLVTFTRQSLDPPPEWERCEKLLTRLH 720  
QY 721 VTYEGTIEGNGQGMLOVDFAANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTVLDHNE 780  
Db 721 VTYEGTIEGNGQGMLOVDFAANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTVLDHNE 780  
QY 781 CLIIITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRTEIIVDAIDALHFRRYLDQFVPE 840  
Db 781 CLIIITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRTEIIVDAIDALHFRRYLDQFVPE 840  
QY 841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
Db 841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
QY 901 FTFGDSELMDIYSMHTFLTERKLTIVGEVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 960  
Db 901 FTFGDSELMDIYSMHTFLTERKLTIVGEVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 960  
QY 961 VESCTQTTNQPGORTGA 977  
Db 961 VESCTQTTNQPGORTGA 977

RESULT 3  
US-09-511-507-2  
; Sequence 2, Application US/095111507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-2

Query Match 100.0%; Score 5184; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MSAGPGCEPCTKRPRWDAATAATSPPAASDARSFPGRRVLDSDKAPVQFRVPPSSGCGAL 60  
Db 1 MSAGPGCEPCTKRPRWDAATAATSPPAASDARSFPGRRVLDSDKAPVQFRVPPSSGCGAL 60  
  
QY 61 GRAGQHRGSATSLVFQKQITTSWMDTKGKTVESLSHSEKNNNTREESMMSSVQKDNFY 120  
Db 61 GRAGQHRGSATSLVFQKQITTSWMDTKGKTVESLSHSEKNNNTREESMMSSVQKDNFY 120  
  
QY 121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHOTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
Db 121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHOTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
  
QY 181 EQFSNANVDQSSPKDDHSDTNSEESRDNQOFLTHVKLANAKQTMEDEQGREARSHQKCGK 240  
Db 181 EQFSNANVDQSSPKDDHSDTNSEESRDNQOFLTHVKLANAKQTMEDEQGREARSHQKCGK 240  
  
QY 241 ACHPAEACAGCQCEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPFFEKESE 300  
Db 241 ACHPAEACAGCQCEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPFFEKESE 300  
  
QY 301 PESPMVDVNSKNSCQDSEADETSPGFDEQEDSSSAQTANKPSRFPQPREADTELKRSSA 360  
Db 301 PESPMVDVNSKNSCQDSEADETSPGFDEQEDSSSAQTANKPSRFPQPREADTELKRSSA 360  
  
QY 361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
Db 361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
  
QY 421 AEDKRKEQCEMKHQRTERKIPKYPHLSPDKKWLGTPTEEMRRMPCGIRLPLRPSAN 480  
Db 421 AEDKRKEQCEMKHQRTERKIPKYPHLSPDKKWLGTPTEEMRRMPCGIRLPLRPSAN 480  
  
QY 481 HTVTIRVDLLRIGEVKPPFTHFKDLWDNKHVKMPCSEQNLYPVEDENGERRAAGSRWELI 540  
Db 481 HTVTIRVDLLRIGEVKPPFTHFKDLWDNKHVKMPCSEQNLYPVEDENGERRAAGSRWELI 540  
  
QY 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLVEEAQAHLYSILPDMVK 600  
Db 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLVEEAQAHLYSILPDMVK 600  
  
QY 601 IALCLPNICTQPIPLLKQKMHSTMTSQQEIASLLANAFCTFPPRRNAKMKSEYSSYPDI 660  
Db 601 IALCLPNICTQPIPLLKQKMHSTMTSQQEIASLLANAFCTFPPRRNAKMKSEYSSYPDI 660

QY 661 NENRLEGRSSRKPEKLTILFCYFERRVTEKPTGLVTFTRQSLDEDFPEWERCEKLLTRLH 720  
Db 661 NENRLEGRSSRKPEKLTILFCYFERRVTEKPTGLVTFTRQSLDEDFPEWERCEKLLTRLH 720  
  
QY 721 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTVLDHNE 780  
Db 721 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTVLDHNE 780  
  
QY 781 CLIIITGEQYSEYTGAEYRWARSHEDRSERDDWQRRRTTEIVAIDALHFRRYLDQFVPE 840  
Db 781 CLIIITGEQYSEYTGAEYRWARSHEDRSERDDWQRRRTTEIVAIDALHFRRYLDQFVPE 840  
  
QY 841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
Db 841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
  
QY 901 FTFGDSELMRDIYSMHTFLTERKLTIVGEVYKLLRLRYNEECRNCTPGDDIKLYPFIYHA 960  
Db 901 FTFGDSELMRDIYSMHTFLTERKLTIVGEVYKLLRLRYNEECRNCTPGDDIKLYPFIYHA 960  
  
QY 961 VESCTQTNNQPGQRTGA 977  
Db 961 VESCTQTNNQPGQRTGA 977

RESULT 4  
US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

Query Match 89.4%; Score 4635.5; DB 4; Length 976;  
Best Local Similarity 89.5%; Pred. No. 0;  
Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
  
QY 1 MSAGPGCEPCTKRPRWDAATAATSPPAASDARSFPGRRVLDSDKAPVQFRVPPSSGCGAL 60  
Db 1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPGRRVLDPKDAHVQFRVPPSPACVP 59  
  
QY 61 GRAGQHRGSATSLVFQKQITTSWMDTKGKTVESLSHSEKNNNTREESMMSSVQKDNFY 120  
Db 60 GQAGQHRGSATSLVFQKQITTSWMDTKGKTAESESLDSKENNTRIESMMSSVQKDNFY 119  
  
QY 121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHOTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
Db 120 QHNVEKLVNSQLSLDKSLTEKSTQYLNQHTAAMCKWQNEGKHTEQLESEPTVTLVP 179  
  
QY 181 EQFSNANVDQSSPKDDHSDTNSEESRDNQOFLTHVKLANAKQTMEDEQGREARSHQKCGK 240  
Db 180 EQFSNANIDRSPQNDHSDTDSEENRDNQOFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239  
  
QY 241 ACHPAEACAGCQCEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPFFEKESE 300  
Db 240 SCHPGEDCASCCQDEIDVVPKSPSLSDVGSEVGTGSKNDNKLIRQESCLGNSPPFFEKESE 299

301 PESPMVDNSKNSCODSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELKRESSA 360  
300 PESPMVDNSKNSCODSEADEETSPGFDEQEDSSSAQTANKPSRFQARDADIEFRKRYST 359  
361 KGGEIRLHFQFEGGESRAGMNDVNKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
360 KGGEVRLHFQFEGGESRTGMNDLNKLPNGNISLNVECRNSKHGRKDSKITDHLMLPK 419  
421 AEDKRKEQCEMKHQRTERKIPKYPHLSPKKWLGTPIEEMRMRCGIRLPLRPSAN 480  
420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPKKWLGTPIEEMRMRCGIRLPLRPSAN 479  
481 HTVTIRVDLLRIGEVKPPFTFHKDLWKNKHVMPCSEONLYPVEDENGERRAAGSRWELI 540  
480 HTVTIRVDLLRIGEVKPPFTFHKDLWKNKHVMPCSEONLYPVEDENGERTAGSRWELI 539  
541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEQAHLYSILPDMVK 600  
540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEQAHLYSILPDMVK 599  
601 IALCLPNICTQPIPLKQKMNHSITMSQEQIASLLANAFCTFPRRNKMKSEYSSYPDI 660  
600 IALCLPNICTQPIPLKQKMNHSITMSQEQIASLLANAFCTFPRRNKMKSEYSSYPDI 659  
661 NFNRLFEGRSRKPPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDPEWERCEKLLTRLH 720  
660 NFNRLFEGRSRKPPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDPEWERCEKPLTRLH 719  
721 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLLFTEVLDHNE 780  
720 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNE 779  
781 CLIIITGTEQYSEVTGYAETRWARSHERDQWRRRTTEIIVDAIDALHFRYLDQFVPE 840  
780 CLIIITGTEQYSEVTGYAETRWARSHERDQWRRRTTEIIVDAIDALHFRYLDQFVPE 839  
841 KIRRELKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
840 KMRRELKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 899  
901 FTFGDSELMDIYSMHIFLTERKLTVGVDVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 960  
900 FTFGDSELMDIYSMHIFLTERKLTVGVDVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 959  
961 VESCTQTTNPGQRTG 976  
960 VESCAETADHSGQRTG 975

RESULT 5  
US-09-511-477-4  
; Patent No. 6337202  
; Sequence 4, Application US/09511477  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:

US-09-511-477-4  
Query Match 89.4%; Score 4635.5; DB 4; Length 976;  
Best Local Similarity 89.5%; Pred. No. 0;  
Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
QY 1 MSAGPGCEPCTKRPRWDAATSPPAASDARSFPGRQRRVLDSDAPVQFRVPPSSSGCAL 60  
Db 1 MNAGPGCEPCTKATRNGAATTS-PAASDARSFPGRQRRVLDSDAPVQFRVPPSSPACVP 59  
QY 61 GRAGQHRGSATSLVFKQKTIITSMWDTKGIKTVESESLHSENNNTRRESMSSVQKDNFY 120  
Db 60 GQAGQHRGSATSLVFKQKTIITSMWDTKGIKTAESESLSKENNNTRIESMMSSVQKDNFY 119  
QY 121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHTAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
Db 120 QHNVEKLVNVSQSLDKSLTEKSTQYLNHQHTAAMCKWQNEGKHTEQLLESEPQTTLVP 179  
QY 181 EQFSNANVDQSSPKDDHSDTNSERSDNQQLTHVVKLANAKQTMEDDEQGREARSHQKCGK 240  
Db 180 EQFSNANIDRSPQDDHSDTNSERNQQLTHVVKLANAKQTMEDDEHAREAKSHQKCSK 239  
QY 241 ACHPAEACAGCQOEETDVVSESLSDTGSDDVGTGLKNANRLNRQESSLGNSPPFKEKESE 300  
Db 240 SCHPGEDCASCQOEIDVVPKPSLSDVGSDDVGTGSKNDNKLIRQESCLGNSPPFKEKESE 299  
QY 301 PESPMVDNSKNSCODSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELKRESSA 360  
Db 300 PESPMVDNSKNSCODSEADEETSPGFDEQEDSSSAQTANKPSRFQARDADIEFRKRYST 359  
QY 361 KGGEIRLHFQFEGGESRAGMNDVNKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
Db 360 KGGEVRLHFQFEGGESRTGMNDLNKLPNGNISLNVECRNSKHGRKDSKITDHLMLPK 419  
QY 421 AEDKRKEQCEMKHQRTERKIPKYPHLSPKKWLGTPIEEMRMRCGIRLPLRPSAN 480  
Db 420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPKKWLGTPIEEMRMRCGIRLPLRPSAN 479  
QY 481 HTVTIRVDLLRIGEVKPPFTFHKDLWKNKHVMPCSEONLYPVEDENGERRAAGSRWELI 540  
Db 480 HTVTIRVDLLRIGEVKPPFTFHKDLWKNKHVMPCSEONLYPVEDENGERTAGSRWELI 539  
QY 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEQAHLYSILPDMVK 600  
Db 540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEQAHLYSILPDMVK 599  
QY 601 IALCLPNICTQPIPLKQKMNHSITMSQEQIASLLANAFCTFPRRNKMKSEYSSYPDI 660  
Db 600 IALCLPNICTQPIPLKQKMNHSITMSQEQIASLLANAFCTFPRRNKMKSEYSSYPDI 659  
QY 661 NFNRLFEGRSRKPPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDPEWERCEKLLTRLH 720  
Db 660 NFNRLFEGRSRKPPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDPEWERCEKPLTRLH 719  
QY 721 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLLFTEVLDHNE 780  
Db 720 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNE 779  
QY 781 CLIIITGTEQYSEVTGYAETRWARSHERDQWRRRTTEIIVDAIDALHFRYLDQFVPE 840  
Db 780 CLIIITGTEQYSEVTGYAETRWARSHERDQWRRRTTEIIVDAIDALHFRYLDQFVPE 839  
QY 841 KIRRELKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
Db 840 KMRRELKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 899  
QY 901 FTFGDSELMDIYSMHIFLTERKLTVGVDVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 960  
Db 900 FTFGDSELMDIYSMHIFLTERKLTVGVDVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 959  
QY 961 VESCTQTTNPGQRTG 976  
Db 960 VESCAETADHSGQRTG 975









Db 952 VESSAETDMPGQKAG 967

RESULT 9

US-09-511-507-6

; Sequence 6, Application US/09511507

; Patent No. 6395543

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AME, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/511,507

; CURRENT FILING DATE: 2000-02-23

; PRIOR APPLICATION NUMBER: 09/302,812

; PRIOR FILING DATE: 1999-04-30

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6

; LENGTH: 968

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

US-09-511-507-6

Query Match 83.6%; Score 4332.5; DB 4; Length 968;

Best Local Similarity 84.0%; Pred. No. 0;

Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRRVLDSDAPVQFRVPPSSGCGAL 60

Db 1 MSAGPGWEPTK-ARWGAAGTSAPTASDSRSFPGRRVLDSDAPVQFRVPPSSPACVS 59

QY 61 GRAGQHRGATSLSVFKOKTITSWMDTKGIKTVESESLHSEKNNNTREESMSSVQKDNFY 120

Db 60 GRAGPHRGNATSFVFKOKTITTWMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY 116

QY 121 QHNMEKLENSQLGFDKSPVEKGTQYLKQHQTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Db 117 PHKVEKLENVQNLNLDKSPTEKSSQVILNQQTASVCKWQNEGKHAEQLLASEPPAGTLP 176

QY 181 EQFSNANVDQSSPKDDHSDTNSSESRDNQQFLTHVKLANAKQTMEDQGREARSHQKCGK 240

Db 177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233

QY 241 ACHPAEACAGCQOEETDVVSESLSDTGSDDVGTGLKNANRLNRQESSLGNSPPFKESE 300

Db 234 SRQSVKDTGTCQOEVDVLPESPLSDVGAEDIGTGPKNDKNLTGQESSLGDSPPFKESE 293

QY 301 PESPMVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFPQPREADTELKRSSA 360

Db 294 PESPMVDNSRNSCQDSEADEETSPVDFEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352

QY 361 KGGEIRLHFQEGGESRAGMNDVNARPGSTSLNVECRNSKQHGKDSKITDHFMRVPK 420

Db 353 KGSEVRLHFQEP-GENNAGTSDLNAPKPSGSSSLNVECRSSKQHGKDSKITDHFMRISK 411

QY 421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSDPKCKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Db 412 SEDRRKEQCEVNRHQRTERKIPKYIPPNLPPEKCKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

QY 481 HTVTIRVDLLRIGEVKPPFTHFVKDLWDNKVMPKPCSEONLYPVEDENGERRAAGSRWELI 540

Db 472 HTVTIRVDLLRAGEVPKPPFTTHYKOLWDNKVMPKPCSEONLYPVEDENGERTAGSRWELI 531

QY 541 QTALLNRLTRPQNLKDAILKYNVAYSCKKWDFTALIDFWDKVLLEEAQAHLYSILPDMVK 600

Db 532 QTALLNKFTRPQNLKDAILKYNVAYSCKKWDFTALIDFWDKVLLEEAQAHLYSILPDMVK 591

QY 601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAMKSEYSSYPDI 660

Db 592 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAMKSEYSSYPDI 651

QY 661 NFNRLFEGRSRKPPEKLTLCYFRRVTEKKPTGLVTFTQTSLEDFPEWERCEKLLTRLH 720

Db 652 NFNRLFEGRSRKPPEKLTLCYFRRVTEKKPTGLVTFTQTSLEDFPEWERCEKPLTRLH 711

QY 721 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTTEVLDHNE 780

Db 712 VTYEGTIEGNGRMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTTEVLDHNE 771

QY 781 CLIIITGEQYSEYTGAEYRWARSHEDRSERDDWQRRRTTEIIVADALHFRFVLDQFVPE 840

Db 772 CLIIITGEQYSEYTGAEYRWARSHEDGSEKDDWQRRCTEIVADALHFRFVLDQFVPE 831

QY 841 KIRRELNKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQIILAAAVERDVVY 900

Db 832 KVRRELNKAYCGFLRPGVSPSENLSAVATGNWCGAFGGDARLKALIQIILAAAVERDVVY 891

QY 901 FTFGDSELMRDIYSMHTFLTERKLTGVEVYKLLRYNNEECRNCSTPGPDIKLYPFIYHA 960

Db 892 FTFGDSELMRDIYSMHTFLTERKLDVGKVKLLRLRYNNEECRNCSTPGPDIKLYPFIYHA 951

QY 961 VESCTQTNTNPGQRTG 976

Db 952 VESSAETDMPGQKAG 967

RESULT 10

US-09-302-812-8

; Sequence 8, Application US/09302812B

; Patent No. 6333148

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/302,812B

; CURRENT FILING DATE: 1999-04-30

; EARLIER APPLICATION NUMBER: 60/083,768

; EARLIER FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 8

; LENGTH: 768

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE:

US-09-302-812-8

Query Match 20.5%; Score 1063.5; DB 4; Length 768;

Best Local Similarity 41.5%; Pred. No. 5.9e-90;

Matches 244; Conservative 93; Mismatches 198; Indels 53; Gaps 15;

QY 417 RVPKAEADKRKEQCEMKHQRTERKIPKYIPPHL--SPDKKWLGTPIEEM--RRMPCGIRL 472

Db 45 RMSKSPDGGISIEETEE-----PENLANSLDSDSWRGVSMEDIAHRNRQPFLENL 94

QY 473 PPLRPSANHTVTIRVDLLRIGEVKPPFTHFVKDLWDNKVMPKPCSEONLYPVEDENGERRA 532

Db 95 PFVTAGNLHRVMYQ--LPIRETP-PRPKSPGKWDSEHVRLPKAPESKYPRENPDGSTT 150

QY 533 AGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSCKKWDFTALIDFWDKVLLEEAQAHLIQ 592

Db 151 IDFRWEMIERALLQPIKTCEELQAAIISNTTYRDDQWHFRALHQLLDEELDESETRVFFE 210

QY 593 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 651

Db 211 DLLPPIIRLALRPLDLIQSPVPLLKHHKNASLSLSQQIISCLLANAFLCTFPRRNTLKRK 270

|    |     |  |     |
|----|-----|--|-----|
| QY | 652 | SEYSSYPDINFNRLFEGRRSRKPEKLTLCFYFRRV--TEKK-----PTGLVTFTQRS--L   | 703 |
| Db | 271 | SEYSTFPDINFNRLYQSTGPAVLEKCLKIMHYFRRVCPTERDASNVTGVTTFVRRSGLP    | 330 |
| QY | 704 | EDFPEWERCEKLL--TRLHVTYEGTIEGNGQGMLQVDFANRFVGGVTSAGLVQVEEIRFL   | 761 |
| Db | 331 | EHLIDWSQSAAPLGVDVPLHVDAGETIEDEGILLQVDFANKYLGGLGGVLCVCQVEEIRFV  | 390 |
| QY | 762 | INPELIVSRLFTEVLHDNECLITGTQEYSEYTYGAETYRWARSHERSDDDWQRRTTE      | 821 |
| Db | 391 | ICPELLVGKLFTECLRPFEALVMLGAERYSNYTYAGSFWSGNFEDSTPRDSSGRRQTA     | 450 |
| QY | 822 | IVALIDHFRYLDQFVPEKIRRELNKAYCGFLR-----PGVSSSENLSAVATGNWCGGA     | 875 |
| Db | 451 | IVALIDHFAQSHHQYREDLMERELNKAYIGFVHVMVTPPPG-----VATGNWCGGA       | 502 |
| QY | 876 | FGGDARLKALIQILAAAVAERDVVYTFGDSSELMRDIYSMHTFLTERKLTIVGEVYKLLLR  | 935 |
| Db | 503 | FGGDSYLKALLQLMVCQAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS--ILR | 561 |
| QY | 936 | YYNEECRNCSTPGP-----DILKYPFIYHAVESCTQTNTNPQGORTGA               | 977 |
| Db | 562 | YSYRLIKEKSSKEPRENKASKKLYDFI---KBELKKVRDVPGEGASA                | 606 |

RESULT 11  
 US-09-511-477-8  
 ; Sequence 8, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIA2 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 US-09-511-477-8

|    |                       |   |                    |                 |             |          |
|----|-----------------------|---|--------------------|-----------------|-------------|----------|
|    | Query Match           | 20.5%;  | Score 1063.5;      | DB 4;           | Length 768; |          |
|    | Best Local Similarity | 41.5%;  | Pred. No. 5.9e-90; |                 |             |          |
|    | Matches 244;          | Conservative  | 93;                | Mismatches 198; | Indels 53;  | Gaps 15; |
| QY | 417                   | RVPKAEKRKEQCCEMKHQRTERKIPKYIPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL  | 472                |                 |             |          |
|    |                       | : : : : : :   | :                  | :               | :           | :        |
| Dd | 45                    | RMSKSPDGGISEIETEE-----PENLANGSLDDSWRGVSMEAHHRNRQPFELENL       | 94                 |                 |             |          |
|    |                       | : : : : : :   | :                  | :               | :           | :        |
| QY | 473                   | PPLRPSANHTVTIRVDLLRLRIGVPEKPPPTHFKOLDWNKHVKMPCSEQLYPVEDENGERA | 532                |                 |             |          |
|    |                       | :: :: : : : :   | :                  | :               | :           | :        |
| Dd | 95                    | PPVTAGNLHRVMYQ---LPIRETP-PRPYKSPGKWDSEHVRLPCAPESKYPRENPDGSTT  | 150                |                 |             |          |
|    |                       | : : : : : :   | :                  | :               | :           | :        |
| QY | 533                   | AGSRWELIQALLNRLTRPQNLDAILKNVAYSKKWDFLTALIDFWDKVLEEAQAHLQY     | 592                |                 |             |          |
|    |                       | :     : : : : : :   | :                  | :               | :           | :        |
| Dd | 151                   | IDFRWEMIERALLOPIKTCEELQAAISYNTTYRDQWHFRALHQLLDEELDESETRVFFE   | 210                |                 |             |          |
|    |                       | :     : : : : : :   | :                  | :               | :           | :        |
| QY | 593                   | SILPDMVKIALCLPNICTQPIPLLKQKMNHSTMTSQEQIASILLANAFFCTFPRRNA-KMK | 651                |                 |             |          |
|    |                       | : : : : : :   | :                  | :               | :           | :        |
| Dd | 211                   | DLLPRIRLALRPDLIQSPVPLLKHKNVASLSLSQQISCLLANAFACLTFPFRRTLKRK    | 270                |                 |             |          |
|    |                       | : : : : : :   | :                  | :               | :           | :        |
| QY | 652                   | SEYSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L   | 703                |                 |             |          |
|    |                       | :     : : : : : :   | :                  | :               | :           | :        |

|  |     |  |  |     |
|--|-----|--|--|-----|
| Db   | 271 | SEYSTPDPDINFNRLYQSTGPAVLKCLKIMHYFRRVCP | TERDASNVTGVVTFVRRSGLP                        | 330 |
| QY   | 704 | EDFPEWERCEKLL--TRLHVTYEGTIEGNGQGMLQVDF | FANRFVGGVTSAGLVQEEIRFL                       | 761 |
| Db   | 331 | EHLDWSQSAAPLGVDVPLHVD                  | AEGTIEDEGIGLLQVDFANKYLGGLGGVLGHCVCQEEIRFV    | 390 |
| QY   | 762 | INPELIVSRLLFTEVL                       | DHNECLITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRTTE | 821 |
| Db   | 391 | ICPELLVKGKLFTECLRP                     | FEALVMLGAERYSNYTGAGSFEWSGNFEDSTPRDSSGRRQTA   | 450 |
| QY   | 822 | IWAIDALHFRRYLDQFVPEK                   | IRRELNKAYCGFLR-----PGVSSENLSAVATGNWGC        | 875 |
| Db   | 451 | IWAIDALHFAQSHHQYREDL                   | MERELNKAYIGFVHMVTPPPG-----VATGNWGC           | 502 |
| QY   | 876 | EGGDARLKALIQILAAVAERDVVYTF             | FGDSELMRDIYSMHTFLTERKLTVGEVYKLLLR            | 935 |
| Db   | 503 | FGGDSYLKALLQLMVCAQLGR                  | PLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR      | 561 |
| QY   | 936 | YNEECRCNCSTPGP-----DIKLYPFIYH          | AVESCTQTTNPQGQGTGA                           | 977 |
| Db   | 562 | SYSRLIXBKSSKEPRENKASKKLYDPI---         | KEELKKVRDVPGEASA                             | 606 |
| RESULT 12  |     |  |  |     |
| US-09-511-507-8  |     |  |  |     |
| ; Sequence 8, Application US/09511507  |     |  |  |     |
| ; Patent No. 6395543   |     |  |  |     |
| ; GENERAL INFORMATION:   |     |  |  |     |
| ; APPLICANT: JACOBSON, Myron K.  |     |  |  |     |
| ; APPLICANT: JACOBSON, Elaine L.   |     |  |  |     |
| ; APPLICANT: AME, Jean-Christophe  |     |  |  |     |
| ; APPLICANT: LIN, Winston  |     |  |  |     |
| ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN |     |  |  |     |
| ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV |     |  |  |     |
| ; TITLE OF INVENTION: THEREWITH  |     |  |  |     |
| ; FILE REFERENCE: NIAD 201   |     |  |  |     |
| ; CURRENT APPLICATION NUMBER: US/09/511,507  |     |  |  |     |
| ; CURRENT FILING DATE: 2000-02-23  |     |  |  |     |
| ; PRIOR APPLICATION NUMBER: 09/302,812   |     |  |  |     |
| ; PRIOR FILING DATE: 1999-04-30  |     |  |  |     |
| ; NUMBER OF SEQ ID NOS: 38   |     |  |  |     |
| ; SEQ ID NO 8  |     |  |  |     |
| ; LENGTH: 768  |     |  |  |     |
| ; TYPE: PRT  |     |  |  |     |
| ; ORGANISM: Drosophila melanogaster  |     |  |  |     |
| ; FEATURE:   |     |  |  |     |
| US-09-511-507-8  |     |  |  |     |

|    | Query Match           | 20.5%;   | Score 1063.5;      | DB 4;           | Length 768; |
|----|-----------------------|--|--------------------|-----------------|-------------|
|    | Best Local Similarity | 41.5%;   | Pred. No. 5.9e-90; |                 |             |
|    | Matches 244;          | Conservative   | 93;                | Mismatches 198; | Indels 53;  |
|    |                       |  |                    | Gaps            | 15;         |
| QY | 417                   | RVPKAEKRRKEQCEMKHQRTERKIPKIYPHLL--SPDKKWLGTPIIEM--RRMPRCGIRL   | 472                |                 |             |
| DB | 45                    | RMSKSPDGGISIEIETEE-----PENLANSLDSDRGVSMEDIAHRNRQPFLENL         | 94                 |                 |             |
| QY | 473                   | PPLRPSANHTVTRVDLLRIGEVKPFPTHFKDLWDNKHVKWPCSEQNLYPVEDENGERA     | 532                |                 |             |
| DB | 95                    | PPVTAGNLHRVMYQ---LPIRETP-PRPYKSPGKWDSEHVRLPCAPESKYPRENPDGSTT   | 150                |                 |             |
| QY | 533                   | AGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDF TALIDFWDKVL EEAQAHL YQ | 592                |                 |             |
| DB | 151                   | IDFRWEMIERALLQPIKTCIELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE   | 210                |                 |             |
| QY | 593                   | SILPDMVKIALCLPNICTOPIPLLKQKNHSITMSQEQIASLLANAFCTFPRRNA-KMK     | 651                |                 |             |
| DB | 211                   | DLLPRIIRLALRPLDLIQSPVPLLKHKKNASLSLSQQQISCLLANAFCTFPRRNTLKRK    | 270                |                 |             |
| QY | 652                   | SEYSSYPDINFNRLFEGRSRKP EKLKTLFCYFRRV--TEKK-----PTGLVTFTRQS--L  | 703                |                 |             |
| DB | 271                   | SEYSTFPDINFNRLYQSTGEAVLEKLK CIMHYFRRVCPPTERDASNVPTGWTFVRRSGLP  | 330                |                 |             |
| QY | 704                   | EDFPWECKERCKLL--TRLHVTYEGTIEGNQGQMLQVDFANRFVCGGVT SAGLVQEEIRFL | 761                |                 |             |



Db 331 EHLIDWSQSAAPLGDVPLHVDAGETIEDEGIGLLQVDFANKYLGGLGGVGHGCVQEEIRFV 390  
QY 762 INPELIVSRFLTVELDHNECLIIITGTEQYSEYTGVAETYRWARSHEDRSERDDWQRRTE 821  
Db 391 ICPELLVGKLFTECLRPPEALVMLGAERYSNYTGAGSFEWSGNFEDSTPRDSSGRRQTA 450  
QY 822 IVAIDALHFRYLDQFVPEKIRRELKAYCGFLR-----PGVSSENLSAVATGNWCGGA 875  
Db 451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHVMVTPPPG-----VATGNWCGGA 502  
QY 876 FGGDARLKALIQILAAVAERDVVYFTFGDSSELMRDIYSMHTFLTERKLTGVEVYKLLR 935  
Db 503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561  
QY 936 YNNEECRNCSTPGP-----DIKLYPFIYHVESCTQTNTNPGQRTGA 977  
Db 562 SYSRLIKESSKEPRENKASKKLYDFI---KEELKKVRDVPGEASA 606  
  
RESULT 13  
US-09-302-812-10  
; Sequence 10, Application US/09302812B  
; Patent No. 633148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-302-812-10  
  
Query Match 9.0%; Score 465; DB 4; Length 726;  
Best Local Similarity 24.3%; Pred. No. 3.5e-34;  
Matches 204; Conservative 116; Mismatches 282; Indels 236; Gaps 35;  
  
QY 199 DTNSESRDNQQFLTHVKLANAKQTME-----DEQGREARSHQKCGKACHPAEACAGCQOE 254  
Db 10 DPVTQDEKDYEDYV-GVGFAHQVPTMKRRKLTGHNNTTESKE-----DPEE-----PK 56  
  
QY 255 ETDVVSSEPLSDTGSSED-----VGTGLKNA-----NRLNRQESSLGN 291  
Db 57 SRDVFVSSQSSDESQEDSAENPEIAKEVSENCELTETLKISNIESLDNVTERSEHTLDN 116  
  
QY 292 ---SPPFKESEPESPMDV-----DNSKNSCQSEADEETS PGF--DEQE--- 331  
Db 117 HKSTEPMEEDVNNKSNIDVAINSDDEDELVLLENNKEMRDGEQVQQLSQDLFADDQELIE 176  
  
QY 332 -----DSSSAQTANKPSRFQPREADTELKRSSAKGGEIRLHFQFEGGESR 377  
Db 177 YPGIMKDTTQDITDSEVETAQKMEMIEEADS-----TFVGEDSK 219  
  
QY 378 AGMNDVNAKRPGSTSSLNVECRNSQHGKDKSKITDHFMRVPKAEDKRKEQCEMKHQRT 437  
Db 220 A-----TKVTRTSSSSFLST-----VSTCEAPAKGRARMYQKELE 254  
  
QY 438 RKIPKYIPHLIS--PDKKWLGTPIEEMRMPCRGIRLPPLRPSANHTVTIRVDLLRIGEV 495  
Db 255 KHVIAFTEGNLTLOPD-----LNKVDPRNY-----RYCTI 285

QY 496 PKPFP-THFKDLWDNKH---VKMPCSEQNLYPVEDENGERAAGSRWELIQTALLNRLTRP 551  
Db 286 PN-FPASQGLKREDNRYGPKIVLPQRWREF-----DSRGR----- 320  
  
QY 552 QNLKDAILKYN---VAYSKQWDFTA-----LIDFWD---KVLLEAEAAQHLYQSILPDM 598  
Db 321 ---RDSYFYFKRKLGDYLCYKTKTYGFMFVGLLHNWFEFDPDITYKLPALEMYKEMSEL 377  
  
QY 599 V-----KIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPR 645  
Db 378 VGREEVLEKFARVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCAALVARMFF----- 431  
  
QY 646 RNAKMSEYSSYPDINFNRLFEGRSSRKPEKLTLCFCYFRRVTEKKPTGLVTF--TRQSL 703  
Db 432 --ARPDSPFS-----FCRILSSDKSICVEKLFKFLTYFDKMSMDPPDGAVSFLTKMDK 483  
  
QY 704 EDFPEWERCEKLLTRLHVTY--EGTIEGNGQGMLOVDVANRFVGGGVTSAGLVQEEIRFL 761  
Db 484 DTFNEEWKOKKLSLPEVEFFDEMLIEDTAL-CTQVDFAHEHLGGGVNLHGSGVQEEIRFL 542  
  
QY 762 INPELIVSRFLTVELDHNECLIIITGTEQYSEYTGVAETYRWA-----RSHEDRSB-RDDW 815  
Db 543 MCPMMVMGMLLCEKMKQLEAISIVGAYVFSSYTGCGHTLTKWAELOPNHSRQNTNEFRDRF 602  
  
QY 816 QRRTEIVAIDALHFRR-----YLDQFVPEKIRRELKNAKAYCGFLRPGVSSENLSAVATGN 870  
Db 603 GRLRVETIAIDAILFKGSKLDCQTEQLNKANIREMKKASIGFMSQGPXFTNI-PIVTGW 661  
  
QY 871 WCGAFGGDARLKALIQILAAVAERDVVYFTFGDSSELMRDIYSMHTFLTERKLTGVE 928  
Db 662 WCGAFNGDKPLKFIQVIAAGVADRPLHFCFSGEPELAACKKIEMKQKDVTLGK 719  
  
RESULT 14  
US-09-511-477-10  
; Sequence 10, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-477-10  
  
Query Match 9.0%; Score 465; DB 4; Length 726;  
Best Local Similarity 24.3%; Pred. No. 3.5e-34;  
Matches 204; Conservative 116; Mismatches 282; Indels 236; Gaps 35;  
  
QY 199 DTNSESRDNQQFLTHVKLANAKQTME-----DEQGREARSHQKCGKACHPAEACAGCQOE 254  
Db 10 DPVTQDEKDYEDYV-GVGFAHQVPTMKRRKLTGHNNTTESKE-----DPEE-----PK 56  
  
QY 255 ETDVVSSEPLSDTGSSED-----VGTGLKNA-----NRLNRQESSLGN 291  
Db 57 SRDVFVSSQSSDESQEDSAENPEIAKEVSENCELTETLKISNIESLDNVTERSEHTLDN 116  
  
QY 292 ---SPPFKESEPESPMDV-----DNSKNSCQSEADEETS PGF--DEQE--- 331  
Db 117 HKSTEPMEEDVNNKSNIDVAINSDDEDELVLLENNKEMRDGEQVQQLSQDLFADDQELIE 176



QY 332 -----DSSSAQTANKPSRFOPREADTELKRSSAKGGEIRLHFQFEGGESR 377  
Db 177 YPGIMKDTTTLQDITDSEVETAQKMEMIEETEADS-----TFVGEDSK 219  
QY 378 AGMNDVNAKRPGSTSSLNVECRNSKQHGKDKSKITDHFMVRPKAEDKRKEQCEMKHQRT 437  
Db 220 A-----TKTVRTSSSSFLST-----VSTCEAPAKGRARMYQKELE 254  
QY 438 RKIPKYIPPHLS--PDKKWLGTPIEEMRRMPCGIRLPLRPSANHTVIRVDLLRIGEV 495  
Db 255 KHVIAFTEGNLTLPD-----LNKVDPRNY-----RYCTI 285  
QY 496 PKPFP--THFKDLWDNKH---VKMPCSEQNLVPVEDENGERRAAGSRWELIQTALLNRLTRP 551  
Db 286 PN-FPASQGLREDNRYGPKIVLPQRWREF-----DSRGR----- 320  
QY 552 QNLKDAILKYN---VAYSKWDFTA-----LIDFWD---KVLEEAQAHLYSILPDM 598  
Db 321 ---RDSYFYFKRKLDGYLKCYKTTGYFMFVGLLHNMWFEFDPDITYKLPALEMYKEMSEL 377  
QY 599 V-----KIALCLPNICTQPIPLLKQKMNHSITMSOEQIASLLANAFCTFPR 645  
Db 378 VGREEVLEKFAVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCAALVARMFF----- 431  
QY 646 RNAKMSEYSSYPDINENRLEFGRSSRKPEKLTLCYFRRVTEKKPTGLVTF--TRQSL 703  
Db 432 --ARPDSPFS-----FCRILSSDKSICVEKLTLCYFRRVTEKKPTGLVTF--TRQSL 703  
QY 704 EDFPEWERCEKLTLLRLHVTY--EGTIEGNGQGMQLQVDFANRFVGGGVTSAGLVQEEIRFL 761  
Db 484 DTFNEEWKOKKRLSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGVNLHGVSQVEIRFL 542  
QY 762 INPELIVSRLEFTEVLHDNECLIIITGTEQYSEYTGVAETRYWA-----RSHEDRSE-RDDW 815  
Db 543 MCPPEMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHRSQNTNEFRDRF 602  
QY 816 QRRTEIIVADALHFR-----YLDQFVPEKIRRELNKAYCGFLRPGVSSENLSAVATGN 870  
Db 603 GRLEVETIADAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPFTNI-PIVTGW 661  
QY 871 WCGGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVGE 928  
Db 662 WCGGAFNGDKPLKFIQVIAAGVADRPLHFCFSFGEPELAACKKIIERMKQKDVTLGK 719

RESULT 15  
US-09-511-507-10  
; Sequence 10, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-507-10

Query Match 9.0%; Score 465; DB 4; Length 726;  
Best Local Similarity 24.3%; Pred. No. 3.5e-34;

Matches 204; Conservative 116; Mismatches 282; Indels 236; Gaps 35;  
QY 199 DTNSESRDNQQFLTHVKLANAKQTME---DEQGREARSHQKQKACHPAEACAGCQOE 254  
Db 10 DPVTQDEKDYEDYV-GVGFAHQVPTMKRKLTEHGNTTESKE-----DPEE-----PK 56  
QY 255 ETDVVSSEPLSDTGSDE-----VGTGLKNA-----NRLNRQESSLGN 291  
Db 57 SRDVFVSSQSSDESQEDSAENPEIAKEVSENCELTETLKISNIESLDNVTERSEHTLDN 116  
QY 292 ---SPPEKESEPEPMDV-----DNSKNSCQDSEADEETSPGF--DEQE--- 331  
Db 117 HKSTEPMEEDVNNKSNIIDVAINSDDEDELVLLENKEMRDGEQVQQLSQDLFADDQELIE 176  
QY 332 -----DSSSAQTANKPSRFOPREADTELKRSSAKGGEIRLHFQFEGGESR 377  
Db 177 YPGIMKDTTTLQDITDSEVETAQKMEMIEETEADS-----TFVGEDSK 219  
QY 378 AGMNDVNAKRPGSTSSLNVECRNSKQHGKDKSKITDHFMVRPKAEDKRKEQCEMKHQRT 437  
Db 220 A-----TKTVRTSSSSFLST-----VSTCEAPAKGRARMYQKELE 254  
QY 438 RKIPKYIPPHLS--PDKKWLGTPIEEMRRMPCGIRLPLRPSANHTVIRVDLLRIGEV 495  
Db 255 KHVIAFTEGNLTLPD-----LNKVDPRNY-----RYCTI 285  
QY 496 PKPFP--THFKDLWDNKH---VKMPCSEQNLVPVEDENGERRAAGSRWELIQTALLNRLTRP 551  
Db 286 PN-FPASQGLREDNRYGPKIVLPQRWREF-----DSRGR----- 320  
QY 552 QNLKDAILKYN---VAYSKWDFTA-----LIDFWD---KVLEEAQAHLYSILPDM 598  
Db 321 ---RDSYFYFKRKLDGYLKCYKTTGYFMFVGLLHNMWFEFDPDITYKLPALEMYKEMSEL 377  
QY 599 V-----KIALCLPNICTQPIPLLKQKMNHSITMSOEQIASLLANAFCTFPR 645  
Db 378 VGREEVLEKFAVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCAALVARMFF----- 431  
QY 646 RNAKMSEYSSYPDINENRLEFGRSSRKPEKLTLCYFRRVTEKKPTGLVTF--TRQSL 703  
Db 432 --ARPDSPFS-----FCRILSSDKSICVEKLTLCYFRRVTEKKPTGLVTF--TRQSL 703  
QY 704 EDFPEWERCEKLTLLRLHVTY--EGTIEGNGQGMQLQVDFANRFVGGGVTSAGLVQEEIRFL 761  
Db 484 DTFNEEWKOKKRLSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGVNLHGVSQVEIRFL 542  
QY 762 INPELIVSRLEFTEVLHDNECLIIITGTEQYSEYTGVAETRYWA-----RSHEDRSE-RDDW 815  
Db 543 MCPPEMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHRSQNTNEFRDRF 602  
QY 816 QRRTEIIVADALHFR-----YLDQFVPEKIRRELNKAYCGFLRPGVSSENLSAVATGN 870  
Db 603 GRLEVETIADAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPFTNI-PIVTGW 661  
QY 871 WCGGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVGE 928  
Db 662 WCGGAFNGDKPLKFIQVIAAGVADRPLHFCFSFGEPELAACKKIIERMKQKDVTLGK 719

Search completed: May 26, 2004, 18:49:20  
Job time : 35.3065 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:16:17 ; Search time 59.0768 Seconds  
(without alignments)  
13253.917 Million cell updates/sec

Title: US-09-302-812-1  
Perfect score: 7242  
Sequence: 1 accggaagtgaacgaagcc.....aaattttcattaacaaaaa 4070

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2 1/USPTO spool\_p/6333148/runat\_26052004\_150054\_5689/app\_query.fasta\_1.12437  
-DB=PIR\_78 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148@CGN\_1\_1\_298 @runat\_26052004\_150054\_5689 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID     | Description        |
|------------|-------|-------------|--------|--------|--------------------|
| 1          | 522   | 7.2         | 997    | B84726 | probable poly(ADP- |
| 2          | 465   | 6.4         | 726    | T21138 | hypothetical prote |
| 3          | 331   | 4.6         | 364    | A84726 | probable poly(ADP- |
| 4          | 185.5 | 2.6         | 1320   | S57113 | BUD4 protein - yea |
| 5          | 179   | 2.5         | 1213   | A58198 | serine/proline-ric |
| 6          | 175.5 | 2.4         | 3942   | T42730 | Bassoon protein -  |
| 7          | 175.5 | 2.4         | 5327   | T13564 | microtubule-associ |
| 8          | 175   | 2.4         | 1359   | T34036 | hypothetical prote |
| 9          | 174.5 | 2.4         | 1403   | A47328 | natural killer cel |
| 10         | 173   | 2.4         | 1210   | T39410 | AF-4 protein, spli |
| 11         | 170.5 | 2.4         | 1200   | A46194 | neurofilament prot |
| 12         | 170.5 | 2.4         | 1507   | B47328 | natural killer cel |
| 13         | 170.5 | 2.4         | 1576   | T03277 | pol protein - yeas |
| 14         | 170.5 | 2.4         | 6642   | T29757 | protein UNC-89 - C |

|    |       |     |      |   |        |                     |
|----|-------|-----|------|---|--------|---------------------|
| 15 | 170   | 2.3 | 1071 | 2 | E85343 | hypothetical prote  |
| 16 | 168   | 2.3 | 778  | 2 | A35970 | erythrocyte-bindin  |
| 17 | 168   | 2.3 | 1115 | 2 | T19137 | hypothetical prote  |
| 18 | 167.5 | 2.3 | 646  | 1 | S15901 | chromogranin B pre  |
| 19 | 165.5 | 2.3 | 1320 | 2 | JC5630 | TCOF1 protein - mo  |
| 20 | 165   | 2.3 | 2248 | 2 | A35938 | profilaggrin - hum  |
| 21 | 162.5 | 2.2 | 7160 | 2 | T27935 | hypothetical prote  |
| 22 | 162   | 2.2 | 1231 | 2 | T18532 | serine/threonine pr |
| 23 | 161.5 | 2.2 | 943  | 2 | S54493 | hypothetical prote  |
| 24 | 161.5 | 2.2 | 1535 | 2 | T49042 | hypothetical prote  |
| 25 | 161   | 2.2 | 2022 | 2 | A59256 | myosin-IXb [simila  |
| 26 | 160   | 2.2 | 2176 | 2 | T13806 | toucan gene protei  |
| 27 | 158.5 | 2.2 | 1016 | 2 | T19006 | ankyrin related pr  |
| 28 | 158   | 2.2 | 489  | 2 | A45988 | dentin matrix acid  |
| 29 | 158   | 2.2 | 1424 | 2 | T03851 | thyroid hormone re  |
| 30 | 157.5 | 2.2 | 1791 | 2 | T02345 | hypothetical prote  |
| 31 | 157   | 2.2 | 1192 | 2 | A71623 | probable secreted   |
| 32 | 157   | 2.2 | 3122 | 2 | T17202 | DNA-directed DNA p  |
| 33 | 156.5 | 2.2 | 1269 | 2 | F84730 | probable myosin he  |
| 34 | 156   | 2.2 | 1222 | 2 | T22490 | hypothetical prote  |
| 35 | 156   | 2.2 | 1280 | 2 | T00365 | hypothetical prote  |
| 36 | 156   | 2.2 | 1804 | 2 | T34518 | nestin - golden ha  |
| 37 | 156   | 2.2 | 3147 | 2 | T18674 | hypothetical prote  |
| 38 | 155.5 | 2.1 | 522  | 2 | C96608 | AAS surface protei  |
| 39 | 155.5 | 2.1 | 1463 | 2 | T30290 | nucleolin homolog   |
| 40 | 154.5 | 2.1 | 611  | 2 | T06458 | hypothetical prote  |
| 41 | 154.5 | 2.1 | 1877 | 2 | T21861 | SRP40 protein - ye  |
| 42 | 154   | 2.1 | 406  | 2 | S38170 | neurofilament trip  |
| 43 | 154   | 2.1 | 854  | 2 | S02003 | probable GTPase-ac  |
| 44 | 154   | 2.1 | 1822 | 2 | T14106 | NF-180 - sea lampr  |
| 45 | 153.5 | 2.1 | 1110 | 2 | I51116 |                     |

ALIGNMENTS

RESULT 1

B84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84726  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-997 <STO>  
A;Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g31870  
A;Map position: 2

|   |          |  |      |
|---|----------|--|------|
| Alignment Scores:                         | 6.62e-28 | Length:  | 997  |
| Pred. No.:                                | 522.00   | Matches:   | 136  |
| Score:                                    | 45.88%   | Conservative:  | 59   |
| Percent Similarity:                       | 32.00%   | Mismatches:  | 116  |
| Best Local Similarity:                    | 7.21%    | Indels:  | 114  |
| Query Match:                              | 2        | Gaps:  | 12   |
| DB:                                       |          |  |      |
| US-09-302-812-1 (1-4070) x B84726 (1-997) |          |  |      |
| QY  | 1989     | TGGGATAAGGTACTAGAGAAGCAGAAGCTCAACTGTATCAGTCCTTCCTGAT         | 2048 |
| Db  | 79       | PheAspGluLeuIleAspGluLysGluSerLysArgTrpPheAspGluIleIleProAla | 98   |
| QY  | 2049     | ATGGTGAAAATTGCACCTCTGTCTGCCAAATATT                           | 2081 |
| Db  | 99       | LeuAlaSerLeuLeuLeuGlnPheProSerLeuGluValHisPheGlnAsnAlaAsp    | 118  |

QY 2082 -----TGTACCAGCCAAATACCACTCTGAAACAGAAAGATGAATCATTCC 2126  
Db 119 AsnIleValSerGlyIleLysThrGlyLeuArgLeuLeuAsnSerGlnGlnAlaGlyIle 138  
QY 2127 ATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAAATGCTTTTCTTCTGCAGCTTT 2186  
Db 139 ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158  
QY 2187 CCACGACCAATGCCAAGATGAATCAGAGTATTCAGTTATCCAGATATTAACCTTCAAT 2246  
Db 159 ProAspAsnArgGlyAlaLys-----HisLeuProValIleAsnPheAsp 174  
QY 2247 CGG-----TTGTTGAAGGACGTTTCATCAAGGAAACCAGAGAAGCTTAAACGCTCTTC 2300  
Db 175 HisLeuSerLeuTyriIleSerTyriSerGlnSerGlnGluSerLysIleArgCysIleMet 194  
QY 2301 TGCTACTTTAGAAAGATCACAGAGAAAAAACCCACTGGGTGGTGACATTACAAAGACAG 2360  
Db 195 HisTyriPheGluArgPheCysSerCysValProIleGlyIleValSerPheGluArgLys 214  
QY 2361 -----AGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAAACCCTGACTCGACTG 2414  
Db 215 IleThrAlaAlaProAspAlaAspPheTrpSerLysSerAspValSerLeu----- 231  
QY 2415 CATGTCACTTACGAAGGTACCATAGAGGAAACGGCCAGGGCATGCTACAGGTGGATTTT 2474  
Db 232 -----TyrGln-----ProAspAsnAlaLeuGluValAspPhe 242  
QY 2475 GCAAAACCGTTTCGTTGGAGGTGGTGTAAACAGTGCAGGACTTTGTGCAAGAAAGAAATCCGC 2534  
Db 243 AlaAsnLysTyriLeuGlyGlyGlySerLeuSerArgGlyCysValGlnGluIleArg 262  
QY 2535 TTTTAAATCAACCCCTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAAT 2594  
Db 263 PheMetIleAsnProGluLeuIleAlaGlyMetLeuPheLeuProArgMetAspAspAsn 282  
QY 2595 GAATGCTTATCATCAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAACA 2654  
Db 283 GluAlaIleGluIleValGlyAlaGluArgPheSerCysTyriThrGlyTyriAlaSerSer 302  
QY 2655 TACCGTGGCCCGGAGCCATGAAGACAGAGCGGAAAGGACGACTGGCAGAGCGGCACG 2714  
Db 303 PheArgPheAlaGlyGluTyriIleAspLysLysAlaMetAspPropheLysArgArg 322  
QY 2715 ACTGAGATCGTCGCATCGACGCCCTC-----CACTCAGACGCTAC 2756  
Db 323 ThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLysAspIle 342  
QY 2757 CTCGACCACTTTGTGCCGAGAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTT 2816  
Db 343 Cys-----LeuLeuArgGluIleAsnLysAlaLeuCysGlyPhe 355  
QY 2817 CTTCGT----- 2822  
Db 356 LeuAsnCysSerLysAlaTrpGluHisGlnAsnIlePheMetAspGluGlyAspAsnGlu 375  
QY 2822 ----- 2822  
Db 376 IleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrThrAlaSer 395  
QY 2822 ----- 2822  
Db 396 HisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnAsnLeuIle 415  
QY 2823 -----CCTGGAGTTTCTTCAGAGAACCTG-----TCTGCAGTGGCTACA 2861  
Db 416 ArgAspPheTyriValGluGlyValAspAsnGluAspHisGluAspAspGlyValAlaThr 435  
QY 2862 GGAACCTGGGCTGTGGTGCCTTTGGGGTGATGCTAGACTAAAAGCCTTAATACAGATC 2921  
Db 436 GlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIleGlnTrp 455

QY 2922 CTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTT---TATTTCACCTTTGGGGACTCAGAA 2978  
Db 456 LeuAlaAlaSerGlnThrArgArgProPheIleSerTyriThrPheGly---ValGlu 474  
QY 2979 CTGATGAGAGACATT 2993  
Db 475 AlaLeuArgAsnLeu 479  
RESULT 2  
T21138  
hypothetical protein F20C5.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C;Accession: T21138  
R;Matthews, P.  
submitted to the EMBL Data Library, December 1995  
A;Reference number: Z19381  
A;Accession: T21138  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-726 <WIL>  
A;Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1  
A;Experimental source: clone F20C5  
C;Genetics:  
A;Gene: CESP:F20C5.1  
A;Map position: 4  
A;Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1  
C;Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1  
Alignment Scores:  
Pred. No.: 5.29e-24 Length: 726  
Score: 465.00 Matches: 204  
Percent Similarity: 38.19% Conservative: 116  
Best local Similarity: 24.34% Mismatches: 282  
Query Match: 6.42% Indels: 236  
DB: 2 Gaps: 35  
US-09-302-812-1 (1-4070) x T21138 (1-726)  
QY 852 GACACAAATAGTGAGGAGGTAGAGATAATCAGCAGTTTTTTGACACATGTAAAGCTTGC 911  
Db 10 AspProValThrGlnAspGluLysAspTyriGluAspTyriVal---GlyValGlyPheAla 28  
QY 912 AATGCAAAAGCAGACGATGAA-----GATGAACAGGCGACAGAAAGCCAGAAAGC 959  
Db 29 HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThrGluSer 48  
QY 960 CACCAGAACTGTGGCAAGGCTTGCATCTGCGAAGCCCTGTGCGAGGTGTGACGAGGAG 1019  
Db 49 LysGlu-----AspProGluGlu-----ProLys 56  
QY 1020 GAGACAGACGTGTGTCGAGAGCCCCCTTGTGCGACACTGGCTCTGAGGAT----- 1070  
Db 57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76  
QY 1071 -----GTTGGTACTGGACTGAAAAATGCC----- 1094  
Db 77 AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96  
QY 1095 -----AACAGATTGAATAGACAAAGTAGTAGTCTAGGAAAT 1130  
Db 97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrIleuAspAsn 116  
QY 1131 -----TCTCCTCCATTGTAGAAAGAAAGTGAACCTGAGTCACCAATGGATGA--- 1178  
Db 117 HisLysSerThrGluProMetGluAspValAsnAsnLysSerAsnIleAspValAla 136  
QY 1179 -----GATAATTCCAAAAATAGTTGTCTCAGGAT 1205  
Db 137 IleAsnSerAspGluAspGluLeuValLeuGluAsnAsnLysGluMetArgAsp 156  
QY 1206 TCAGAAGCAGATGAAGACACAAGTCCAGGTTTT-----GATGAACAGGAA----- 1250

Db 157 GlyGluGlnValGlnGlnLeuSerGlnAspLeuPheAlaAspAspGlnGluLeuIleGlu 176  
QY 1251 -----GATAGCAGTTCTGCTCAA 1268  
Db 177 TyrProGlyIleMetLysAspThrThrGlnLeuAspIleThrAspSerGluValGlu 196  
QY 1269 ACAGCAATAAACCCTTCAAGGTTCCAAACCAAGAGAAGCTGACACTGAGTTGAGGAACGG 1328  
Db 197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer----- 211  
QY 1329 TCCTCTGCTAAGGAGGTGAGATTTCGATTACATTTCCAATTGAAGGAGGAGAGTCTGA 1388  
Db 212 -----ThrPheValGlyGluAspSerLys 219  
QY 1389 GCTGGAATGATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCTGAATGTAGAG 1448  
Db 220 Ala-----ThrLysThrValArgThrSerSerSerPheLeuSerThr--- 234  
QY 1449 TGCAGAAATCTAAGCAACATGGGAGAAAGGATTCTAAATCACAGATCATTTTCATGAGA 1508  
Db 234 ----- 234  
QY 1509 GTGCCCAAAGCAGAGGACAAAAGAAAGAAACAATGTGAATGAACATCAAGAAGACAGAA 1568  
Db 235 ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254  
QY 1569 AGGAAGATCCCTAAATACATTCCACCTCACCTTTCT-----CCAGATAAGAAATGGCTT 1622  
Db 255 LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp----- 270  
QY 1623 GGAACCTCTATTGAGGAGATGAGGAGAAATGCCAAGGTGTGGGATCCGGCTGCCCTTGG 1682  
Db 271 -----LeuAsnLysVal 274  
QY 1683 AGACCATCTGCCAATCACACAGTGACTATTTCGGGTAGATCTTTTGGCAATAGGAGAAGTT 1742  
Db 275 AspProAspArgAsnTyr-----ArgTyrCysThrIle 285  
QY 1743 CCTAAACCTTTCCCA---ACACATTTTAAAGATTGTGGGACAAACAGCAT----- 1790  
Db 286 ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304  
QY 1791 GTTAAGATGCCTTGTTCAGAACAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA 1850  
Db 305 IleValLeuProGlnArgTyrArgGluPhe-----AspSerArgGlyArgArg 320  
QY 1851 GCTGCAGGCAGCGGTGGGAACCTCATTCAGACTGCACCTTCTCAACAGGCTCACTCGGCC 1910  
Db 320 ----- 320  
QY 1911 CAGAACCTGAAGGATGCTATTCTGAAGTACAAT-----GTGGCATATTCTAAGAAA 1961  
Db 321 -----ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337  
QY 1962 TGGGACTTTTACAGCT-----TTGATTGATTTCTGGGAT----- 1994  
Db 338 TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTyrGluPheAsp 357  
QY 1995 ---AAGTACTAGAAGAACGAGAGCTCAACACTGTATCATCAGTCCATCTTGCCTGATATG 2051  
Db 358 ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeu 377  
QY 2052 GTG-----AAAATTGCACCTCTGTCTG 2072  
Db 378 ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAlaLysThrAla 397  
QY 2073 CCAAAATATTGTACCCAGCCAAATACCACCTCCTGAAACAGAAAGATGAATCATCTCCATCACA 2132  
Db 398 GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluserAlaThr 416  
QY 2133 ATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGA 2192  
Db 417 LeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431

QY 2193 CGCAATGCCAAGATGAATCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGTTG 2252  
Db 432 -----AlaArgProAspSerProPheSer-----PheCysArgIle 443  
QY 2253 TTTGAAGGACGTTTCATCAAGGAAACAGAGAGCTTAAACGCTCTTCTGCTACTTTTGA 2312  
Db 444 LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463  
QY 2313 AGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTC-----ACAAGACAGAGTCTT 2366  
Db 464 LysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483  
QY 2367 GAAGATTTCAGAGTGGGAAAGATGTGAAAAAATCCTGACTCGACTGCATGTCACTTAC 2426  
Db 484 AspThrPheAsnGluGluTyrLysAspLysLysLeuArgSerLeuProGluValGluPhe 503  
QY 2427 -----GAAGGTACCATAGAGGAAACGGCCAGGCGCATGCTACAGGTGGATTTCGAAAAC 2480  
Db 504 PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522  
QY 2481 CGTTTCGTTGGAGGTGGTGTAAACAGTCAGGACGCTTGTGCAAGAAATCCGCTTTTAA 2540  
Db 523 GluHisLeuGlyGlyValLeuAsnHisGlySerValGlnGluIleArgPheLeu 542  
QY 2541 ATCAACCTGAGTTGATTGTTTTCAGGGCTCTTCACTGAGGTCTGGATCACAATGAATGT 2600  
Db 543 MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuAla 562  
QY 2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCGGAAACATACCGC 2660  
Db 563 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLys 582  
QY 2661 TGGGCC-----CGGAGCCATGAAGACAGAGCGGAA---AGGACGACTGG 2702  
Db 583 TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGlnPheArgAspArgPhe 602  
QY 2703 CAGAGGCGCAGACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGC----- 2753  
Db 603 GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu 622  
QY 2754 -----TACCTCGACACAGTTTGTGCCGAGAAAGATCAGACGGGAGCTTAACAAGCTTAC 2807  
Db 623 AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSer 642  
QY 2808 TGTGATTTCCTCGTCTCGAGTTTCTTCAGAGAACCTGTCTGCTGCTGCTACAGGAAAC 2867  
Db 643 IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTyr 661  
QY 2868 TGGGGCTGTGGTCCCTTTGGGGGTGATGCTAGACTAAAGCCCTTAATACAGATCCTGGCA 2927  
Db 662 TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681  
QY 2928 GCTGCTGTAGCTGAGCGAGACGCTGTTTATTTCACCTTTGGGACTCAGAACTGATGAGA 2987  
Db 682 AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAla 701  
QY 2988 GACATTTACAGCATGCATACATTCTCCTCACTGAGAGGAAACTGACTGTTGGAGAA 3041  
Db 702 LysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyLys 719  
RESULT 3  
A84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: A84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197





Db 290 SerThrSerThrLysSerGluGluSerTyrIleAlaAspTyrLysValThrArgGlnGlu 309  
QY 519 -----GGAATCAAGACAGTGAATCAGAAAGTTTGCATAGTAAAGAAAC----- 563  
Db 310 AspTyrAspThrLysLysLeuHisGlnGluSerGluHisAlaAsnGlnGlnProAlaIle 329  
QY 564 -----AACAAATACAGAGAAGAATCCATGATGAGTCTGTACAAAAAGATAAATTTAT 617  
Db 330 IleProGlnLysAspSerSerGluGluThrPheThrGluLeuAsnAsnGluSerGluPhe 349  
QY 618 CAACATAACATG----- 629  
Db 350 GlnArgAsnPhelysAspGlyGluGluTyrArgIleValGlnHisGluGluSerLeuTyr 369  
QY 630 -----GAAAAATTAGAAAAATGTTTCTCAGCTAGCTTTTGTAT 665  
Db 370 GlyGlnArgThrLysSerProGluGluAsnIleAsnGlySerGluIleGlyValAsp 389  
QY 666 -----AAGTCACCAGTTGAAAAAGGTACACAGTATTGTAAGCAG 704  
Db 390 HisGlyGluAlaAlaGluValAsnGluProLeuAlaLysThrSer-----AlaGluGlu 407  
QY 705 CATCAGACTGGGCTATGTGTAAGTGGCAG-----AATGAAGGGCCACACTCAGAACCGG 758  
Db 408 HisAspLeuSerSerCysGluAspGlnSerValSerGluAlaArgAsnLysAspArg 427  
QY 759 CTTTGTGGAA----- 767  
Db 428 IleGluGluLysGluValGluThrLysAspGluAsnIleGluThrGluLysAspGluSer 447  
QY 768 -----AGTGAACCTCCAGCGGTAACTCTGTGTACCAGAG--- 800  
Db 448 GluTyrHisLysValGluGluAsnGluProGluHisValProLeuLeuProLeu 467  
QY 801 -----CAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCAAGGAT 842  
Db 468 ProArgTyrGluGluIleGlnPheAsnGluProPheIleAspGluAsnAspThrSerAsn 487  
QY 843 GATCACAGTGAC----- 854  
Db 488 AspSerIleAspLeuThrArgSerMetLysProSerAspTyrIleSerIleTyrHisIle 507  
QY 855 -----ACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTGTACA 896  
Db 508 GlnGluGluIleLysSerAsnSerProGluSerIleAlaAsnSerGlnPheSerGln 527  
QY 897 CATGTAAAGCTTGCGAATGCAAAAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGA 956  
Db 528 GlnSerSerIleThrThrAla---SerThrValAspSerLysLysAspAsnGlySerThr 546  
QY 957 AGCCACCAG-----AAGTGTGGCAAGGCTTGCCATCCTCGCAGAAGCC 998  
Db 547 SerPheLysPheLysProArgIleValSerArgSerArgIleTyrAsnProLysSerArg 566  
QY 999 TGTGCAGGGTGTTCAG-----CAGGAGGAGACAGACGTGGTGTCCGAGAGCCCC----- 1046  
Db 567 ValSerSerLeuAsnTyrTyrAspAsnGluAspTyrIleLeuSerAsnSerGluTyrAsn 586  
QY 1047 -----TTGTGGACACTGGCTCTGAGGATGTTGGT 1076  
Db 587 AlaLeuAspProMetArgArgAsnThrLeuIleSerLysArgIleGlnAspAsnIleArg 606  
QY 1077 ACTGGACTGAAAAATGCC-----AACAGATTGAATAGACAAGAA 1115  
Db 607 ThrGlnLysGlyHisAlaProLeuIleArgProSerIleMetLysLeuAsnGlyGluAsp 626  
QY 1116 AGTAGTCTAGGAAATTCCTCCATTGTGAGAAAGAAAGTGAACCTGAGTCACCAATGGAT 1175  
Db 627 SerGlyPheGlnAsn-----HisPheLeuGluValGluGlnProGlnGlu 641  
QY 1176 GTAGATAATCCAAAAATAGTTGTCAGGATTTCAGAACGACAGATGAAGACACAAGTCCAGGT 1235  
Db 642 HisGluAsnIleProLeuSerThrHisLeuSerGluGlnAspIleThrThrAsnValGly 661

QY 1236 TTTGATGAACAGAGAAAGATAGCAGTTCTGCTCAAAACAGCAAATAAACCTTCAAGGTTCCAA 1295  
Db 662 LeuAspGluGlnLysLeuProThrAsnThrGlnAsp----- 673  
QY 1296 CCAAGAGAAAGCTGACACTGAGTTTGAGGAAGCGGTCTCTGCT----- 1337  
Db 674 -----GluAlaGluIleSerIleArgGluIleGluSerAlaGlyAspIleThrPheAsn 691  
QY 1338 AAGGGAGGTGAGATTGATTACATTTCATTTTCAATTTGAAGGAGGAGAGCTGAGCTGGAATG 1397  
Db 692 ArgGlyAspLeuLeuSerLeuSerPheAspGluGluLeuGlyGlnAspPheAlaAsnPhe 711  
QY 1398 AATGATGTG-----AAT 1409  
Db 712 LeuAlaLeuAspHisAspSerThrSerPheAsnHisGlyProAspAspSerSerSer 731  
QY 1410 GCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATCTAAGCAA--- 1466  
Db 732 PheGlnArgAspSerSerLysLysSerPheAsnSerLeuTyrGluSerSerTyrGluLeu 751  
QY 1467 -----CATGGGAGAAAGGATTCTAAAAATCACAGATCATTTCATGAGAGTGCCC 1514  
Db 752 LysProProProSerIleArgLysGlnProIleAlaProAspValLeuGlnLysLeu--- 770  
QY 1515 AAAGCAGAGGACAAAAGAAAGAAACAATGTGAAATGAAACATCAAAAGAACAGAAAGG--- 1571  
Db 771 ---LeuGluSerAspThrLysAspAspAlaAspLeuGluLysIleArgGluGluArgIle 789  
QY 1572 -----AAGATCCCT----- 1580  
Db 790 ThrGluProArgThrGlyLeuGlyIleGlyMetLeuLysThrProValLysAspValSer 809  
QY 1580 ----- 1580  
Db 810 IleAlaLeuAlaAlaSerIleLysGlyTyrGluAlaSerPheSerAspThrAspSerArg 829  
QY 1580 ----- 1580  
Db 830 ProGluGlyMetAsnAsnSerAspAlaIleThrLeuAsnMetPheAspAspPheGluGlu 849  
QY 1581 ---AAATACATTCACCT-----CACCTTCTCCAGATAAGAAATGGCTTGA 1625  
Db 850 AspLysMetThrProSerThrProValArgSerIleSerProIleLysArgHisValSer 869  
QY 1626 ACTCCT----- 1631  
Db 870 SerProPheLysValValLysAlaGlyAsnLysGlnGluAsnAsnGluIleAsnIleLys 889  
QY 1632 -----ATTGAGGAGATGAGGAGAAATGCCAAGGTGTGGATCCGG-----CTG 1673  
Db 890 AlaGluGluGluIleGluProMetThrGlnGlnGluThrAspGlyLeuLysGlnAspIle 909  
QY 1674 CCTCCCTTGAGACCATCTGCCAATCACACAGTCACTATTTCGGGTAGATCTTTCCGAATA 1733  
Db 910 ProProLeuLeuAlaGlnThrLysAspAsnValGluAlaLysGluGluThrIleThrGln 929  
QY 1734 GGAGAAAGTTCCTAACCTTTCCCAACACATTTTAAAGAT----- 1772  
Db 930 LeuGluGluProGlnAspValGluGlnGluPheProAspMetGlyThrLeuTyrLeuSer 949  
QY 1773 -----TTGTGGGACAAACAGCATGTTAAGATGCTTGTGTTCA 1808  
Db 950 IleLysAlaIleSerThrLeuAlaLeuTyrGlyThrLys-----Ser 963  
QY 1809 GAACAAAACCTTGACCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGACCGCGGTGG 1868  
Db 964 HisArgAlaThrTyrAlaIleValPheAspAsnGlyGluAsnValValGlnThrProTyr 983  
QY 1869 GAACACTCATT----- 1877  
Db 984 GluSerLeuProTyrAspGlyAsnIleArgIleAsnLysGluPheGluLeuProIleAsp 1003

QY 1878 -----CAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCT 1928  
Db PheLysGlyLysThrGluThrSerAlaSerSerGluArgAspSerTyrLysLysCys 1023  
QY 1929 ATTCTGAAGTACAATGTGGCAATCTTAAG---AAATGGGACTTTACAGCTTTTGATGAT 1985  
Db ValIleThrLeuLysCysLysTyrGluLysProArgHisGluLeuValGluIleValAsp 1043  
QY 1986 -----TTCTGGGATAAGGTA 2000  
Db LysValProValGlyLysSerPhePheGlyLysThrLysTyrLysPheGluLysLysTyr 1063  
QY 2001 CTAGAAGAAGCAGAAGCTCAACACTTGTATCATCAGTCCATCTTGGCTGATATGGTGAAT 2060  
Db ValGlnLysLysProLysGlnAspGluTyrAspTyrLeuPheAlaGlnAspGlySerPhe 1083  
QY 2061 GCACCTGTCTGCCAAATATTTGTACCCAGCCCAATACCACCTCTGAAACAGAGATGAAT 2120  
Db AlaArgCys----- 1086  
QY 2121 CATTCATCAACAATGTACAGGAACAGATGCCAGTCTTTTAGTAATGCTTTCTTCTGC 2180  
Db ---GluIleGluIleAsnGluGlu-----PheLeuLysAsnValAlaPheAsn 1101  
QY 2181 ACGTTTCCAGCAGCAATGCCAAGATGAAATCAGAGTATTCAGTTATCCAGT----- 2234  
Db Thr---SerHisMetHisTyrAsnMetIleAsnLysTrpSerArgIleAlaAspLysIle 1120  
QY 2235 ATTAACCTCAATCGTTGTTTGAAGGACGTTTCATCAAGGAACCCAGAGAAGCTTAAACG 2294  
Db HisGlySerLysArgLeuTyrGlu---LeuProArgLysAlaProHisLysValAlaSer 1139  
QY 2295 CTC-----TTCTGCTACTTTAGAAGAGTACACAGAGAAAAACCCACTGGTGGTG 2345  
Db LeuAspValGluAlaCysPheLeuGluArgThrSer----- 1151  
QY 2346 ACATTACAAGACAGAGTCTTGAAGATTTTCCA---GAGTGGGAAGATGTGAAAACTC 2402  
Db ---AlaPheGluGlnPheProLysGlnPheSerLeuValAsnLysIle 1166  
QY 2403 CTGACTCGACTG-----CATGTCACCTTACGAAGGTACCATAGAAGGAACGCG 2450  
Db ValSerLysTyrLysLeuGlnGlnAsnIleTyrLysGluGlyTyrLeuLeuGlnAspGly 1186  
QY 2451 CAGGGCATGTACAGGTGGATTTTGCAACCGTTTCGTT-----GGAGGTGGTGA 2501  
Db ---GlyAspLeuLysGlyLysIleGluAsnArgPhePheLysLeuHisGlySerGlnLeu 1205  
QY 2502 ACCAGTGCAGACTTGTGCAAGAAGAAATCCGCTTTTAAATCAACCTGAGTTGTT 2561  
Db SerGlyTyrHisGluIleSerArgLysAlaLysIleAspIleAsnLeuLysValThr 1225  
QY 2562 -----TCACGCTCTTCACTGAG 2579  
Db LysValLeuArgAsnGluAspIleGlnAlaAspAsnGlyGlyGlnArgAsnPheThrAsp 1245  
QY 2580 GTGCTGGATCAATGAATGTCTTATCATCAGGTACTGAGCAGTACAGTGAATACACA 2639  
Db TrpValLeuPheAsnGluCysPheGlnLeuValPheAspAsp----- 1259  
QY 2640 GGCTATGCCGAACATACCGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGACGAC 2699  
Db GlyGluArgIleThrPheAsnAlaGluCysSerAsnGluGluLysSer-----Asp 1276  
QY 2700 TGGCAGAGGGCGCAGCTGAGATCGTCCGCTCGACGCGCCCTCCAC-----TTCAGA 2750  
Db TrpTyrAsnLysLeuGlnGluValValGluLeuAsnValPheHisGlnProTrpValLys 1296  
QY 2751 CGCTACCTCGACCACTTTGTGCCCGAGAGATCAGACGG 2789  
Db LysTyrCysGluLysLeuAlaGluGluGluLysThrArg 1309

A58198  
serine/proline-rich FEL protein, splice form 1 - human  
N;Alternate names: AF-4 protein  
C;Species: Homo sapiens (man)  
C;Date: 02-Jul-1996 #sequence\_revision 25-Oct-1996 #text\_change 01-Dec-2000  
C;Accession: A58198; 152572  
R;Morrissey, J.J.M.  
Blood 81, 1124-1131, 1993  
A;Title: A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias.  
A;Reference number: 152572; MUID:93184301; PMID:8443374  
A;Accession: A58198  
A;Status: translated from GB/EMBL/DBBJ  
A;Molecule type: mRNA  
A;Residues: 1-1213 <MOR>  
A;Cross-references: GB:L25050; NID:g407323  
A;Accession: 152572  
A;Status: translated from GB/EMBL/DBBJ  
A;Molecule type: mRNA  
A;Residues: 'MAQS', 13-1213 <MO2>  
A;Cross-references: GB:L25050; NID:g407323; PIDN:AAA36642.1; PID:g407324  
C;Comment: This protein is one of several to form chimeric fusion proteins with the trit  
C;Genetics:  
A;Gene: GDB:MLLT2; AF-4; FEL  
A;Cross-references: GDB:136792; OMIM:159557  
A;Map position: 4q21-4q21  
C;Keywords: alternative splicing  
  
Alignment Scores:  
Pred. No.: 0.000337 Length: 1213  
Score: 179.00 Matches: 238  
Percent Similarity: 32.72% Conservative: 137  
Best Local Similarity: 20.77% Mismatches: 449  
Query Match: 2.47% Indels: 322  
DB: 2 Gaps: 57

US-09-302-812-1 (1-4070) x A58198 (1-1213)

QY 79 CGTCTGGATAGGCCTGGTTCGGGAGGCTGTCTCAGAGCAGGAGTGCAGAAAGCAGTCAGCG 138  
Db 170 ArgLeuGlyGlnGluGlyPheGly-----SerSerHisHisLysLysGlyAspArgArg 187  
QY 139 GCAGAGGGGCATGGTCCCG-----GAGGCACCGAGAGGGGGCGCAGTCCG 186  
Db 188 AlaAspGlyAspHisCysAlaSerValThrAspSerAlaProGluArgGluLeuSerPro 207  
QY 187 -----TCCCTCCAGGGTTAGT-----AATGAGGCTCTA 216  
Db 208 LeuIleSerLeuProSerProValProProLeuSerProIleHisSerAsnGlnGlnThr 227  
QY 217 CGCCGGCTGGCCCGGAGACTCAGTCTCGGGTCCCGATGAGTCCG----- 267  
Db 228 LeuProArgThrGlnGlySerSerLysValHisGlySerSerAsnSerLysGlyTyr 247  
QY 268 GCCCGGCTGTGAGCCCT-----GCACCAAGC 294  
Db 248 CysProAlaLysSerProLysAspLeuAlaValLysValHisAspLysGluThrProGln 267  
QY 295 GAC-----CCCGCTGGGACCGCTGCAACTTCTCCGCGCGCGCTCGGACG 342  
Db 268 AspSerLeuValAlaProAlaGlnProProSerGlnThrPhe---ProProProSerLeu 286  
QY 343 CCCGAGCTTCCCGGAGCAGGAGCGCGTCTCGATTCCCAAGGACGC-----TCCGG 396  
Db 287 ProSerLysSerValAla-MetGlnGlnLysProThrAlaTyrValArgProMetAspG 306  
QY 397 TGCAGTTACGGTCCCGCGTCTCGTCAGGCTCGCCCTCGCGCGCGCGG----- 448  
Db 306 YGlnAspGlnAlaProSerGluSerProGluLeuLysProLeuProGluAspTyrArgG 326  
QY 449 -----ACAGCACCGAGGCGCGCCCTCTCTTGTGTTTTCACAGAAAGACTATAACAGTT 504  
Db 326 nGlnThrPheGluLys-ThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuL 346



QY 505 GGATGGACACTAAAGGAATCAAGACAGTGTGAATCAGAAAAGTTTGCATAGTAAAGAA----- 560  
Db 346 ysMetProSerGlnSerValGluGlnThrThrTyrSerAsnGluValHisCysValGluGlu 366  
QY 561 -----AACACAATA 570  
Db 366 leLeuLysGluMetThrHisSerTirpProProLeuThrAlaIleHisThrProSert 386  
QY 571 CAAGAGAAGAAATCCATGATGAGTCTGTACAAAAGATAACTTTTATCAACATAACATGG 630  
Db 386 hrAlaGluProSerLysPheProPheProThrLysAspSer-----GlnHisValSers 404  
QY 631 AAAAAATTAGAAAATGTTTCTCAGCTAGGTTTGTATAAGTCCACAGTTGAAAAAGGTACAC 690  
Db 404 erValThrGlnAsnGlnLysGln-----TyrAspThrSer-----SerLysThrH 419  
QY 691 AGTATTTGAAGCAGCATCAGACTCGGGCTATGTGTAAAG-----TGGCAGA 735  
Db 419 isSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspLeuGlnLeuSerAsps 439  
QY 736 ATGAAGGGCCACACTCAGAACGGCTTTTGGAA-----AGTGAACCTCCA- 779  
Db 439 erGluAspSerAspSerGluGlnThrProGluLysProProSerSerSerAlaProProS 459  
QY 780 --GCGGTAACTCTCGTACCAGAGCAGTTTCAGTAATGCTAAT----- 818  
Db 459 erAlaProGlnSerLeuProGluProValAlaSerAlaHisSerSerSerAlaGluSerG 479  
QY 819 -----GTCGATCAGTCGTCCCAAGGATGATCAGAGTGACACAAATAGTGAGGAGA 870  
Db 479 luSerThrSerAspSerAspSerSerSerAspSerGluSerGluSerSerSerAsps 499  
QY 871 GTAGAGATAATCAGCAGTTTTGTACACATGTAAAGCTTGCGAATGCCAAGCAGACGATGG 930  
Db 499 erGluGluAsnGlu-----ProLeuGluThrProAlaProGluProG 513  
QY 931 AAGATGAACAGGGCAGAGAAGCCAGAACCCAGAGAGTGTGGCAAGGCTTGCCATCCTG 990  
Db 513 luProProThrThrAsnLysTrpGlnLeuAspAsnTrpLeuThrLysValSerGlnProA 533  
QY 991 CAGAAAGCCTGTGCGGTGTCAGCAGGAGGAGCAGACGTGGTGTCGAGAGCCCTTGT 1050  
Db 533 laAlaProProGluGly-----ProArgSerThrGluProProArgArgHisProGluS 551  
QY 1051 CGGACACTGGCTCTGAGCATGTTGGTACTGGACTGAAAAATGCCAACAGATTGAATAGAC 1110  
Db 551 er---LysGlySerSerAspSerAlaThrSer-----G 561  
QY 1111 AAGAAAGTAGTCTAGGAAATTCCTCCTCAATTTGAGAAGAAAGTGAA---CCTGAGTCAC 1167  
Db 561 lnGluHisSerGluSerLysAspProProProLysSerSerSerLysAlaProArgAlap 581  
QY 1168 CAATGGATGTAGATAAT---TCCAAAAATAGTTGTCAAGATTTCAGAAAGCAGATGAAGAGA 1224  
Db 581 roProGluAlaProHisProGlyLysArgSerCysGlnLysSerProAla----- 597  
QY 1225 CAAGTCCAGGTTTGTATGAACAGGAAGATAGCAGTTCGTCTCAACAGCAAAATAAACCTT 1284  
Db 598 -----GlnGlnGluProProGlnArgGlnThrValGlyThrLysG 611  
QY 1285 CAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTTGAGGAAGCGGTCTCTGCTAAGGGAG 1344  
Db 611 lnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSerLeuGlnGlyG 631  
QY 1345 GTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATGAATGATG 1404  
Db 631 lyArg-----GluProGlyLeuLeuProTyrGlySerArgAspG 644  
QY 1405 TGAATGCCAAA---CGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTCTA 1461  
Db 644 lnThrSerLysAspLysPro-----LysValL 653  
QY 1462 AGCAACATGGGAGA---AAGGANTCTAAAAATCACAGATCATTTTCATGAGAGTGCCCAAG 1518

Db 653 ysThrLysGlyArgProArgAlaAlaAlaSerAsnGluProLysProAlaValProProS 673  
QY 1519 CAGAGGACAAAAAGAAAGAACATGT----- 1544  
Db 673 erSerGluLysLysLysHisLysSerSerLeuProAlaProSerLysAlaLeuSerGlyP 693  
QY 1545 --GAAATGAAACATCAAGAAACAGAAAGGAAGATCCCTAAATACATT----- 1589  
Db 693 roGluProAlaLysAspAsnValGluAspArgThrProGluHisPheAlaLeuValProL 713  
QY 1590 -----CCACCTCACCTTTTCTCCAGATAAGAAATGGCTTGA----- 1625  
Db 713 euThrGluSerGlnGlyProProHisSerGlySerGlySerArgThrSerGlyCysArgG 733  
QY 1626 -----ACTCCTATTGAGGAGATGAGGAGAATGCCAAGGTGTGGATC----- 1667  
Db 733 lnAlaValValGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspT 753  
QY 1668 -----CGGCTGCCTCCCTTGAGA-----CCATCTGCCAATCACACAGTGACTATTTC 1713  
Db 753 hrLysLeuLeuSerProLeuArgAspThrProProProGlnSerLeuMetValLysIleT 773  
QY 1714 GGGTAGATCTTTTGCGAATAGGAGAGTTCCTAAACCTTTTCCCAACACATTTTAAAGATT 1773  
Db 773 hrLeuAspLeuLeu-----SerArgIleProGlnPro----- 783  
QY 1774 TGTGGACAAACAAGCATGTTAAGATGCTTGTTCAGAAACAAACTTGTACCCTGTG---- 1829  
Db 784 -----ProGlyLysGlySerArgGlnArgLysAlaGluAspLysGlnProProAlaGlyL 802  
QY 1830 -----GAAGATGAGAATGGTGAGCGAGCTGCAGGCAGGCAGGCC 1863  
Db 802 ysLysHisSerSerSerSerSerHisLysGluSerSerLysThrLysProSerArgP 862  
QY 1864 GGTGGGAACCTCATTCAGACTGCACCTTCTCAACAGGCTCCTCGGCCCCAGAACCTGAAGG 1923  
Db 822 ysGlyGluAlaGluArgAspCysAspAsnLysLysIleArgLeuGluLysGluIleLysS 842  
QY 1924 ATGCTATTCTGAAGTACAATGTGGCATAT----- 1952  
Db 842 erGlnSerSerSerSerSerSerHisLysGluSerSerLysThrLysProSerArgP 862  
QY 1953 -----TCTAAGAAATGGGACTTTACAGCTTTTGATTGATTCTTGGGATAAGG 1998  
Db 862 roSerSerGlnSerSerLysLys-----GluM 871  
QY 1999 TACTAGAAGAGCAGAAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAATA 2058  
Db 871 etLeuProProProProValSerSerSerSerGlnLysProAlaLysProAlaLeuLysA 891  
QY 2059 TTGCACTCTGTCTGCCAAATATTGTATCCCAAGCAATACCACCTCCTGAAA-----CAGA 2112  
Db 891 rgSerArgArgGluAlaAspThrCysGlyGlnAspProProLysValProAlaValProA 911  
QY 2113 AGATGAATCATTCATCACAATGTTCACAGGAACAG----- 2147  
Db 911 rgValAsnHisLysAspSerSerIleProLysGlnArgValGluGlyLysGlySerA 931  
QY 2148 -----ATTGCCAGTCTTTTAGCTAATGCTTTC---TTCTGCA 2181  
Db 931 rgSerSerSerAlaAspLysGlySerSerGlyAspThrAlaAsnProPheProValProS 951  
QY 2182 CGTTTCCACGACGCAATGCCAAGATGAATCAGAGTATTCAGTTATCCAGATATTAAC 2241  
Db 951 erLeuProAsnGlyAsnSerLys-----ProGlyLysProGlnValLysP 966  
QY 2242 TCAATCGGTTTGTGAAGGACGTTTCATCAAGAAACCAGAGAAGCTTAAACGCTCTTCT 2301  
Db 966 heAspLysGln-GlnAlaAspLeuHisMetArgGluAlaLysLysMetLysGlnLysAla 985  
QY 2302 GCTACTTTTAGAAGAGTCCAGAGAAAAAACCCACTGGGTTGGTGACATTCCACAGACAGA 2361  
:::





Db 805 TrpGlyArgGlnArgGluGlnGlnAspThr---AlaGluSerSerAspAspPheGlySer 823  
 QY 888 TTTTGTGACACATGTAAGCTTGCGAAT--- 914  
 Db 824 GlnLeuArgHisAspTyrValGluAspSerSerGluGlyLeuSerProLeuProPro 843  
 QY 915 -----GCAAGCAGACGATGGAAGATGAACAG----- 941  
 Db 844 GlnProProAlaArgAlaAspMetThrAspGluGluPheMetArgArgGlnIleLeuGlu 863  
 QY 942 -----GGCAGAGAAGCC 953  
 Db 864 MetSerAlaGluGluAspAsnLeuGluGluAspThrAlaValSerGlyArgGlyLeu 883  
 QY 954 AGAAGCCAC-----CAGAAAGTGTC----- 974  
 Db 884 AlaLysHisSerAlaGlnLysAlaSerAlaArgProArgProGluSerSerGlnGluPro 903  
 QY 975 AAGGCTTGCCATCCTGCAGAAAGCCTGTGCAGGGGTGTGCAGCAGGAGGACACACGTTGGTG 1034  
 Db 904 LysArgArgLeuProHisAsnAlaThrThrGlyTyrGluGluLeuLeuSerGluAlaGly 923  
 QY 1035 TCCGAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACTGGAATAATGCC 1094  
 Db 924 ProAlaGluProThrAspSerSerGlyAlaLeuGlnGlyLeuArgArgPheLysThr 943  
 QY 1095 AACAGATTGAATAGA-----CAAGAAAGTAGTCTAGGAAATTTCTCTCCA 1139  
 Db 944 IleGluLeuAsnSerThrGlySerTyrGlyHisGluLeuAspLeuGlyGlnGlyProAsp 963  
 QY 1140 TTTGAGAAAGAAAGTGAACCTGAGTCACCAATGATGTAGATAATTCCTCAAAATAGTTGT 1199  
 Db 964 ProAsnLeuAspArgGluProGlu-----LeuGluMetGluSerLeuThrGlySerPro 981  
 QY 1200 CAGGATTCAGAGCAGATGAAGACACAAGT-----CCAGGTTTGTAT 1241  
 Db 982 GluAspArgSerArgGlyGluHisSerSerThrLeuProAlaSerThrProSerTyrThr 1001  
 QY 1242 GAA-----CAGGAAGATAGCAGTTCTGTCTCAAAACA 1271  
 Db 1002 SerGlyThrSerProThrSerLeuSerSerLeuGluGluAspSerAspSerSer----- 1019  
 QY 1272 GCAAAATAAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAAGCGGTCC 1331  
 Db 1020 -----ProSerArgArgGlnArgLeuGluGluAlaLysGlnGlnArgLysAlaArg 1036  
 QY 1332 TCTGCTAAGGAGGTGAGATTCCGATTACATTTCCAAATTTGAAGGAGGAGAGTCGAGCT 1391  
 Db 1037 HisArgSerHisGlyProLeuLeuProThrIleGlu----- 1048  
 QY 1392 GGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTPAGCCTGAATGTAGAGTGC 1451  
 Db 1049 -----AspSerSerGluGluGluGluLeu 1056  
 QY 1452 AGAAATCTAAGCAACATGGGAGAAAGGATTCTAAATACAGAT---CATTTTCATGAGA 1508  
 Db 1057 ArgGluGluGluGluLeuLeuArgGluGluGlnGluLysMetArgGluValGluGlnArg 1076  
 QY 1509 GTG---CCCAAAGCAGAGGACAAAAGAAAGAACAAATGTGAATGAACATCAAGAGAACA 1565  
 Db 1077 IleArgSerThrAlaArgLysThrArgArgAspLysGluGluLeuArgAlaGlnArg--- 1095  
 QY 1566 GAAAGGAGATCCCTAAATACATTTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGA 1625  
 Db 1096 ArgArgGluArgSerLysThrProProSerAsnLeuSerProIleGluAsp-----Ala 1113  
 QY 1626 ACTCCTATTGAGGAGATGAGGAGAAATGCCAAGGTGTGGATCCGGCTGCCTCCCTTGAGA 1685  
 Db 1114 SerProThrGluGluLeuArgGln-----AlaAlaGluMetGluGluLeuHis 1129  
 QY 1686 CCATCTGCCAATCACACAGTGACTATTCCGGTAGATCTTTTCCGAATAGGAGAAGTTCT 1745

Db 1130 ArgSerSer----- 1132  
 QY 1746 AAACCTTTCCCAACACACATTTTAAAGATTGTGGACAACAAGCATGTTAAGATGCTTGT 1805  
 Db 1133 -----Cys 1133  
 QY 1806 TCAGAAACAAACTTGTACCTGTGGAGAGATGAGAATGGTGAGCGAGCTGCAGGACGCCG 1865  
 Db 1134 SerGluTyrSerProSerProSerLeuAspSerGluAlaGluThrLeuAspGlyGly--- 1152  
 QY 1866 TGGGAACCTCATTCAGACTGCACCTTCTCAACAGGCTCCTCGGCCCCAGAACCTGAAGGAT 1925  
 Db 1153 -----ProThrArgLeuTyrLys 1158  
 QY 1926 GCTATTCTGAACATAATGTG---GCATATTCTAAGAAATGGGACTTTTACA----- 1973  
 Db 1159 SerGlySerGluTyrAsnLeuProAlaPheMetSerLeuTyrSerProThrGluThrPro 1178  
 QY 1974 -----GCTTTGATTGATTTCTGGGATAAGGTACTAGAAGAAGCAGAGCTCAACAC 2024  
 Db 1179 SerGlySerSerThrThrProSerSerGlyArgProLeuLysSerAlaGluGluAlaTyr 1198  
 QY 2025 TTGTATCAGTCCATCTGCTGATATGTTGATAAATTCACACTCTGTCTGCCAAATATTGT 2084  
 Db 1199 -----GluAspMetMetArgLysAla----- 1205  
 QY 2085 ACCCAGCCAATACCACTCTGAAACAGAGATGAATCATTCATCAATGTACAGGAA 2144  
 Db 1206 -----GluMetLeuGlnArgGlnGlnGly 1213  
 QY 2145 CAGATTGCCAGTCTTTTA-----GCTAATGCTTTCTTCTGCACGTTTCCACGACGC 2195  
 Db 1214 GlnValAlaGlyAlaArgGlyProHisGlyGlyProSerGlnProThrGlyProArgSer 1233  
 QY 2196 AATGCCAAGATGAATCAGAGTATTCAGATTATCCAGATTATTAACCTTCAATCGGTTGTTT 2255  
 Db 1234 GlnGlySerPhe-----GluTyrGlnAspThrGlnAspHisAspTyrGly----- 1248  
 QY 2256 GAAGGACGTTTCATCAGGAACACAGAGAGCTTAAACGCTCTTCTGCTACTTTAGAGA 2315  
 Db 1249 ---GlyArgAlaSerGlnPro----- 1254  
 QY 2316 GTCACAGAGAAAAACCCACTGGGTTG-----GTGACATTCA 2354  
 Db 1255 ValAlaGluSerThrProAlaGlyLeuGlyAlaAlaValTyrGluGluIleLeuGlnThr 1274  
 QY 2355 AGACAGAGTCTTGA-----GATTTTCCAGAGTGG 2384  
 Db 1275 SerGlnSerIleAlaArgMetArgGlnAlaSerSerArgAspLeuGlyPheThrGluAsp 1294  
 QY 2385 GAAAGATGTGAAAACTCCTGACTCGACTGTCATGTCACTTACGAAGGTACCATAGAAGGA 2444  
 Db 1295 LysLysLysGluLysGlnPheLeuAsnAlaGluSerAlaTyrMetAspProMetLysGln 1314  
 QY 2445 AACGGC-----CAGGGCATGCTACAGGTGGATT 2474  
 Db 1315 AsnGlyGlyProLeuThrProGlyThrSerProThrGlnLeuAlaAlaProValSerPhe 1334  
 QY 2475 GCAAAACCGTTTTCGTTGGAGGTGTGTAACCACTGCAGACTTGTCAAGAAGAAATCCGC 2534  
 Db 1335 SerThrSer-----ThrSerSerAspSerSerGlyGlyArgValIleProAspValArg 1352  
 QY 2535 TTTTAAATCAACCCCTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAAT 2594  
 Db 1353 -----ValThrGlnHisPheAlaLysGluProGlnAsp-ProLe 1365  
 QY 2595 GAATGCTTATCATCAGGTAAGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAACA 2654  
 Db 1365 uLysLeuHisSerSerProValSerSerThrLeuThrSerLysGluValGlyMetThrPh 1385  
 QY 2655 TACCGCTGGCCCGGAGCCATGAAGACAGGAGCGAAGGAGGAGCTGGCAGAGGCGCAGC 2714  
 Db 1385 eSerGlnGlyProGlySerProAlaThrThrAlaSerProThr-----ArgGlyTyrMe 1403

```
QY 2715 ACTGAGATCGTCGCATCGACGCGCCTCCACTTCAGACGCTACCTCGACCAAGTTTG 2769
Db 1403 tThrProThrSerProAlaGlySerGluArgSerProSerThrSerThrIle 1421
RESULT 7
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Alignment Scores:
Pred. No.: 0.000841 Length: 5327
Score: 175.50 Matches: 229
Percent Similarity: 35.88% Conservative: 180
Best Local Similarity: 20.09% Mismatches: 484
Query Match: 2.42% Indels: 247
DB: 2 Gaps: 45

US-09-302-812-1 (1-4070) x T13564 (1-5327)
QY 64 AAGAGGGAAGGAGGCGT-----CTGGATAGGCGCTGGTTCGGG 102
Db 1448 LysAlaGluSerArgGluSerIleAlaLysThrHisLysAspGluSerSerLeuAsp 1467
QY 103 AGGCTGTGAGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGCGATGGTGCCG---- 158
Db 1468 LysAlaLysGluGlnGluSerArgGluSerLeuAlaGlu--SerIleLysProGlu 1487
QY 159 --GGAGGCACCGAGGAGGGCG-----CAGTCCGTCC 189
Db 1487 erGlyIleAspGluLysSerAlaLeuAlaSerLysGluAlaSerArgProGluSerValt 1507
QY 190 CTCCCAGGGTTAGTGAATGAGGCTCTACGCCCGGGCTGGCCGAGACTCAGTGTGCGG 249
Db 1507 hrAspLysSerLysGluProSerArgGluSerIleAlaGluSerLeuLysAlaGlu 1527
QY 250 GT-----CCCAGCATGAGTCGGGCGCCCGCTGTGAGCCCTGCACCAGCGACCCCGCT 303
Db 1527 erThrLysAspGluLysSerAlaProProSerLysGlu-----AlaSerArgPro---- 1543
QY 304 GGGACGCGCTGCACTTCTCCGCGCGCGCTCGGACGCCCGGAGCTTCCCGGCAGGC 363
Db 1544 --GlySerValValGluSerValLysAspGluThrGluLysSerLysGluProSerArgA 1563
QY 364 AGAGGCGCGTCTCGAATCCAGGACGCTCCGGTGAGTTTTCAGG----- 407
Db 1563 rgGluSerIleAlaGluSerAlaLysProProIleGluPheArgGluValSerArgPro 1583
QY 408 -----GTCCCGCGCTCCTCGTCAGGCTGCG 432
Db 1583 luSerValIleAspGlyIleLysAspGluSerAlaLysProGluSerArgArgAspSerP 1603
QY 433 CCTGGGCGGGGACAGCAGCAGCGGCGGCGGCGGCTCTCTGTTTTCACAGAGA 492
Db 1603 roLeuAlaSerLysGluAlaSerArgProGluSerValLeuGluSerValLysAspGluP 1623
```

```
QY 493 CTATAACCAAGTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATA 552
Db 1623 roIleLysSerThrGluLysSerArgGluSerValAlaGluSerPheLysAlaAspS 1643
QY 553 GTAAAGAAAACAACAAT-----ACAAGAGAAAGAATCCATGA 588
Db 1643 erThrLysAspGluLysSerProLeuThrSerLysAspIleSerArgProGluSerAlav 1663
QY 589 TGAGTTCTGTACAAAAGATAACTTTTATCAACATAACATGGAA----- 632
Db 1663 alGluAsnValMetAspAlaProPheLysGluThrSerArgProGluSerAlaValGlyS 1683
QY 633 -----AAATTAGAAAATGTTTCTCAGCTAGGTTTGTATAAGTCACCAGTTGAAAAGGT- 686
Db 1683 erMetLysAspGluSerMetSerLysGluProSerArgGluSerValLysAspGlyA 1703
QY 687 --ACACAGTATTGAAGCAGCATCAGACTGCGGCTATGTGTAAAGTGGCAGAAATGAAGGGC 744
Db 1703 laAlaGlnSerArgGluThrSerArgProAlaSerValAlaGluSerAlaLysAspGlyA 1723
QY 745 CACACTCAGAAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGTTACGAGCAGT 804
Db 1723 laAspAspLeuLysGluLeu--SerArgProGluSerThrThr----- 1736
QY 805 TCAGTAATGCTAATGTCGATCAGTCGTCCTCCCAAGGATGATCAGACTGACACAAATAGTG 864
Db 1737 -----GlnSerLysGluAlaGlySerIleLysAspGluLysSerProLeuAlaSerG 1754
QY 865 AGGAGGTAGAGATAATCAGCAGCTTTTGTACACATGTAAAGCTTGGCAATGCAAAGCAGA 924
Db 1754 luGluAlaSerArgProAlaSerValAlaGluSerValLys-----AspGluAlaGluL 1772
QY 925 CGATGGAAGATGAACAGGCGCAGAGAGCC-----AGAAGCCACCAGAAAGTGTGGCA 975
Db 1772 ysSerLysGluGluSerArgGluSerValAlaGluLysSerProLeuProSerLysG 1792
QY 976 AGGCTTGCCATCCTGCAGAAAGCCTGTGCAGGCTGTACAGGAG----- 1019
Db 1792 luAlaSerArgProAlaSerValAlaGluSerIleLysAspGluAlaGluLysSerLysG 1812
QY 1020 -----GAGACAGACGTGCTGCCAGAGCCCTTGTCCGAC----- 1055
Db 1812 luGluSerArgGluSerValAlaGluLysSerProLeuProSerLysGluAlaSerA 1832
QY 1056 -----ACTGGCTCTGAGGATGTTGGTACTGGACTGAAATAATGCCAACAGATTGAATA 1107
Db 1832 rgProAlaSerValAlaGluSerIleLysAspGluAlaGluLysSerLysGluGluSerA 1852
QY 1108 GACAAGAAAGTAGTCTAGGAATTTCTCT---CCATTGAGAAAGAAAGTGAACCTGAGT 1164
Db 1852 rgArgGluSerValAlaGluLysSerProLeuProSerLysGluAlaSerArgProAlaS 1872
QY 1165 CACCAATG-----GATGTAGATAATTCCAAA---AATAGTTGTCCAGATT 1206
Db 1872 erValAlaGluSerIleLysAspGluAlaGluLysSerLysGluGluSerArgGluS 1892
QY 1207 CAGAAGCAGATGAAGAGACAAGTCCAGGTTTGTATGAACAGGAGATAGCAGTTCTGCTC 1266
Db 1892 erValAlaGluLysSerProLeuProSerLysGluAlaSerArgProAlaSerValAlaG 1912
QY 1267 AAACA-----GCAATAAACCTTCAAGTTTCCAAACCAAGAGAGCTGACACTG 1314
Db 1912 luSerIleLysAspGluAlaGluLysSerLysGluGluSerArgGluSerValAlaG 1932
QY 1315 AG-----TTGAGGAAGCGGCTCTCTGCTAAGGAGGTGAGATTTCGATTACATTTC 1365
Db 1932 luLysSerProLeuProSerLysGluAlaSerArgProAlaSerValAlaGluSerIleL 1952
QY 1366 AATTTGAAGGAGGAGAGAGT-----CGAGCTG 1392
Db 1952 ysAspGluAlaGluLysSerLysGluGluSerArgGluSerValAlaGluLysSerP 1972
QY 1393 GAATGAATGATGTGAATGCCAAACGACCTGGAAAGTACTTCT---AGCCTGAATGTAGAT 1449
```



Db 1972 roLeuProSerLysGluAlaSerArgProAlaSerValAlaGluSerIleLysAspGluA 1992  
QY 1450 GCAGAAAATTCTAAGCAACATGGGAGAAAGGATTCT-----AAAATCAGATCATT 1500  
Db 1992 laGluLysSerLysGluGluSerArgGluSerValAlaGluLys-SerProLeuPro 2011  
QY 1501 TCATGAGAGTGCCCAAAGCAGAGGACAAAGAAAGAAACAATGTGAAATGAACATCAAA 1560  
Db 2012 Ser-LysGluAlaSerArgProAlaSerValAlaGluSerIleLysAspGluAlaGluL 2031  
QY 1561 GAACAGAAAGGAAGATCCCTAAATACATTCCACCTCACCTTCTCCAGATAAGAAATGGC 1620  
Db 2031 sSerLysGluGluSer-----ArgArgGluSerValAl 2042  
QY 1621 TTGGAACCTCCTATTGAGGAGATGAGGAGAAATGCCAAGGTGTGGATCCGGCTGCCCT 1680  
Db 2042 aGluLysSerProLeuProSerLysGluAlaSerArgProAlaSerValAlaGluSerI 2062  
QY 1681 TGAGACCATCTGCCAATCACAGTGACTATTTCGGGTAGATCTTTTGGCAATAGGAGAAG 1740  
Db 2062 eiLysAspGluAlaGluLysSerLysGluGluSerArgArgGluSerAlaAlaGluLysSe 2082  
QY 1741 TTCCTAAACCTTTTCCCAACACATTTTAAAGATTTGTGGGACAAACAAGCATGTTAAGATGC 1800  
Db 2082 r-ProLeuPro-----SerLysGluAlaSerArgp 2092  
QY 1801 CTTGTTTCAGAAACAAACTTGTACCTCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCA 1860  
Db 2092 roAlaSerValAlaGlu-----SerValLysAspGluAlaAspLysSerLysGluGluS 2110  
QY 1861 GCCGGTGGGAACCTCATT-----CAGACTGCAC 1887  
Db 2110 erArgArgGluSerMetAlaGluSerGlyLysAlaGlnSerIleLysGlyAspGlnSerP 2130  
QY 1888 TTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGTATT----- 1931  
Db 2130 roLeuLysGluValSerArgProGluSerValAlaGluSerValLysAspAspProvall 2150  
QY 1932 -----CTGAAGTACAATGTGGCATATTCTAAGAAATGGGACTTTACAG 1974  
Db 2150 ysSerLysGluProSerArgArgGluSerValAlaGlySerValThrAlaAsp---SerA 2169  
QY 1975 CTTTGATTGAT-----TTCTGGG 1992  
Db 2169 laArgAspAspGlnSerProLeuGluSerLysGlyAlaSerArgProGluSerValAla 2189  
QY 1993 ATAAGGTACTAGAAGAAGCAGAACTCAACAC-----TTGTATCAGTCCA 2037  
Db 2189 spSerValLysAspGluAlaGluLysGlnGluSerArgArgGluSerLysThrGluSerV 2209  
QY 2038 TCTTGCTGATATGGTGAATTTGCACCTCTGTCTGCCAAATATTGTACCCAGCCATAC 2097  
Db 2209 alleleProLysAlaLysAspAspLysSer---ProLysGluValLeuGlnProVals 2228  
QY 2098 CACTC-----CTGAAACAGAAGATGAATCATTCCTATCAATGTCACAG---GAAC 2145  
Db 2228 erMetThrGluThrIleArgGluAspAlaAspGlnProMetLysProSerGlnAlaGluS 2248  
QY 2146 AGATTGCCAGTCTTTTAGCTAATGCTTTCTCTGACGTTTCCACGA---CGCAATGCCA 2202  
Db 2248 erArgArgGluSerIleAlaGluSerIleLysAlaSerSerProArgAspGluLysSerP 2268  
QY 2203 AGATGAAATCAGAGTATTCCAGTTATCCA-----GATA 2235  
Db 2268 roLeuAlaSerLysGluAlaSerArgProGlySerValAlaGluSerIleLysTyrAspL 2288  
QY 2236 TTAACCTCAATCGGTTGTTGAAGGAC-----GTTTCATCAAGGAAACCCAGAGAAGC 2286  
Db 2288 euAspLysProGlnIle-IleLysAspAspLysSerThrGluHisSerArgArgGluSer 2307  
QY 2287 TTAAACCGCTCTTCTGCTACTTTTAGAGAGTCCAGAGAAAAAACCCACTGGGTTGGTGA 2346

Db 2308 LeuGluAspLysSerAlaValThrSerGluLysSerValSerArgProLeuSerValAla 2327  
QY 2347 CATTCAACAAGACACAGAGTCTTGAAGATTTTCCAGAGTGGGAAAAGATGTGAAAACTCCTGA 2406  
Db 2328 SerAspHisGluAlaAlaValAlaIleGluAspAlaLys----- 2341  
QY 2407 CTCGACTGCATGCTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGTCATGCTACAGG 2466  
Db 2342 -----SerSerIleSerProLysAspLysSerArgProGlyPhe--- 2354  
QY 2467 TGGATTTTGCAACCCGTTTCGTGGAGGTGGTGTAAACCAGTGCAGGACTTGTGCAAGAAG 2526  
Db 2355 ---ValAlaGluThrValSerSerProIleGluGluAlaThrMetGluPheSerLysIle 2373  
QY 2527 AAATCCGCTTTTAAATCAACCCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGG 2586  
Db 2374 GluValValGluLysSerSerLeuAlaLeuSerLeuGlnGlyGlySer-----Gly 2390  
QY 2587 ATCACAATGAATGCTTATCATCACAGTA----- 2616  
Db 2391 GlyLysLeuGlnThrAspSerSerProValAspValAlaGluGlyAspPheSerHisAla 2410  
QY 2617 -----CTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGCTGGGCCC 2667  
Db 2411 ValAlaSerValSerThrValThrProThrLeuThrLysProAlaGluLeuAlaGlnIle 2430  
QY 2668 GGAGCCATGAAGACAGGAGCGAAAGGACGACTGGCAGAGCGGCACGACTGAGATCGTCG 2727  
Db 2431 GlyAlaAlaLysThr---ValSerSerProLeuAspGluAlaLeuArgThrProSerAla 2449  
QY 2728 CCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTCCCGAGAGAGATCAGAC 2787  
Db 2450 ProGluHisIleSerArgAlaAspSer-ProAlaGlu---CysAlaSerGluGlu---- 2466  
QY 2788 GGGAGCTTAACAAGGCTTACTGTGGATTCTTCTGCTGAGATTCTTCAGAGAACCTGT 2847  
Db 2467 -----IleAlaSerGln-AspLysSerProGlnValLeuL 2478  
QY 2848 CTGAGTGGCTACAGGAAACTGGGGCTGTGGTGCTTGGGGGTGATGCTAGACTAAAG 2907  
Db 2478 ysGluSerSerArgProAlaTrpValAlaGluSerLysAspAlaAlaGlnLeuLysS 2498  
QY 2908 CCTTAATACAGATCCTGCAGCTGCTGTAGCTGAGCGAGACGTG 2951  
Db 2498 erSerValGluAspLeuArgSerProValAlaSerThrGluIle 2512

RESULT 8  
T34036  
hypothetical protein B0041.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T34036  
R;Fulton, R.; Wohldmann, P.  
submitted to the EMBL Data Library, April 1997  
A;Description: The sequence of C. elegans cosmid B0041.  
A;Reference number: Z21466  
A;Accession: T34036  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1359 <FUL>  
A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7  
A;Experimental source: strain Bristol N2; clone B0041  
C;Genetics:  
A;Gene: CESP:B0041.7  
A;Map position: 1  
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3  
Alignment Scores:  
Pred. No.: 0.000655 Length: 1359  
Score: 175.00 Matches: 141  
Percent Similarity: 34.73% Conservative: 124  
Best Local Similarity: 18.48% Mismatches: 284  
Query Match: 2.42% Indels: 214





Db 657 ThrLysArgArgIleCysLeuThrGlyThrProLeuGlnAsnLeuMetGluTyrHis 676

QY 2448 -----GGCCAGGGCATGCTA-----CAGGTGGATTTCGCAAAACCGT 2483

Db 677 CysMetValAsnPheValLysProGlyLeuGlyThrLysThrGluPheAlaAsnArg 696

QY 2484 TTCGTTGGAGGTGTAAACCAAGTCAGGACTTGTGCAAGAA-----GAAATC 2531

Db 697 PheVal-----AsnIleLeuAsnArgGlyArgThrLysAspAlaSerProLeuGluVal 714

QY 2532 CGCTTTTAA 2540

Db 715 SerPheMet 717

RESULT 9

A47328

Natural killer cell tumor-recognition protein - human

N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor

C;Species: Homo sapiens (man)

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 24-Sep-1999

C;Accession: A47328

R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.

Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993

A;Title: A cyclophilin-related protein involved in the function of natural killer cells.

A;Reference number: A47328; MUID:93133824; PMID:8421688

A;Accession: A47328

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1403 <AND>

A;Cross-references: GDB:137171; OMIM:161565

A;Experimental source: NK killer cells from adult blood

A;Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIIP:122800)

C;Genetics:

A;Gene: GDB:NKTR

A;Cross-references: GDB:137171; OMIM:161565

A;Map position: 3p23-3p21

C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology

C;Keywords: alternative splicing; lymphocyte

F;60-230/Domain: cyclophilin homology <CYP>

Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 0.000715 | Length:       | 1403 |
| Score:                 | 174.50   | Matches:      | 158  |
| Percent Similarity:    | 33.84%   | Conservative: | 110  |
| Best Local Similarity: | 19.95%   | Mismatches:   | 298  |
| Query Match:           | 2.41%    | Indels:       | 226  |
| DB:                    | 1        | Gaps:         | 32   |

US-09-302-812-1 (1-4070) x A47328 (1-1403)

QY 113 AGCAGAGCTGCAGAAGCAGTCAGCGGAGAGGGGGCATGTGCGGGAGGCACCGAGGA 172

Db 566 ThrHisSerSerArgAspSerTyrArgSerLysSerHisSerGlnSerTyrSerArgGly 585

QY 173 GGGGGCGCAGTCCGTCCCTCCAGGTTAGTGAATGAGGCTTACGCCCGCGGTCGCCG 232

Db 586 -SerSerArgSer-----ArgThrAlaSerLysSerSerHisSerArgSerAr 602

QY 233 GAGACTCACTGCTCGGGTCCAGCATGAGTGGGCCCCCGGCTGTGAGCCCTGCACCAA 292

Db 602 gSerLysSerArg-----SerSerSerLysSerGlyHisArgLysArgAlaSerLy 619

QY 293 GCGACCCCGCTGGGACGCCGCT-----GCAACTTCTCCGCC 328

Db 619 sSerProArgLysThrAlaSerGlnLeuSerGluAsnLysProValLysThrGluProLe 639

QY 329 GGCCGCTCGGACGCCCGGAGC-----TTCCCGCGCAGGCAGAG 367

Db 639 uArgAlaThrMetAlaGlnAsnGluAsnValValGlnProValValAlaGluAsnIl 659

QY 368 GCGCGCTCCTCGAATCCAAGGACGCTCCGGTCAGTTCAGGCTCCCGCGCTCCTCGTCAGG 427

Db 659 eProValIleProLeuSerAspSerProProProSerArgTyrLysProGlyGluLysPr 679

QY 428 CTGC-----GCCCTGGCGCGGGGACAGCACCGGAGGCAGCGCCACCTCTCTGTGTTT 481

Db 679 oTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrThrHisLeuLeuPr 699

QY 482 CAAACAGAGACTATAACCAAGTGGATGGACACTAAAGGAATCAAGACAGTTCGAATCAGA 541

Db 699 oIleGlnSerThrTyr---SerLeuAlaAsnIleLys-----GluThrGlySerSerSe 716

QY 542 AAGTTTGCATAGTAAAGAAACAACAATACAGAGAAGATCCATGATGAGTTCCTGTA-- 599

Db 716 rSerTyrHisLysArgGluLysAsnSerGluSerGlnSerThrTyrSerLysTyrSe 736

QY 600 -----CAAAAAGATAACTTTTA 616

Db 736 rAspArgSerSerGluSerSerProArgSerArgSerArgSerArgSerArgSerTy 756

QY 617 TCAACATAACATGGAAAAAATTAGAAAAATGTT---TCTCAGCTAGGTTTGTGATAAGTCACC 673

Db 756 rSerArgSerTyrThrArgSerArgSerLeuAlaSerHisSerHisSerArgSerArgSerPr 776

QY 674 AGTTGAAAAAGGT-----ACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTG 724

Db 776 oSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArgSerSe 796

QY 725 TAAGTGG-----CAGAAATGAAGGGCCCACTCAGAACGGCTTTTGGAAAAATGA 772

Db 796 rSerTyrThrSerIleSerSerAspAspGlyArgArgAlaLysArgArgLeuArgSerSe 816

QY 773 ACCTCCAGCGGTAACTCTGGTACCAGAGCAGTTCAGTAATGCTAATGTC----- 821

Db 816 rGlyLysLysAsnSerValSerHisLysLysHisSerSerSerSerSerGluLysThrLeuHi 836

QY 822 -----GATCAGTCGTCCCAAGGATGATCAGCTGACACACAA 859

Db 836 sSerLysTyrValLysGlyArgAspArgSerSerCysValArgLysTyrSerGluSerAr 856

QY 860 TAGT-----GAGGAGAGTAGAGATAATCAGCAGTTTTTCACACATGTAAAGCTTCGGA 913

Db 856 gSerSerLeuAspTyrSerSerAspSerGluGln-----SerSerValGlnAlaThrGl 874

QY 914 TGCAAGCAGACGATGGAAGATGAACAGGGCAGAGAGCCAGAGAGCCAGAGAGTGTGG 973

Db 874 nSerAlaGlnGluLysGlu---LysGlnGlyGlnMetGluArgThrHisAsnLysGlnGl 893

QY 974 CAAGGCTTGCCATCCTGTCAGAACCTGTGCAGGCTGTGCAGGAGGAGACAGACGTGGT 1033

Db 893 uLys-----AsnArgGlyGluGluLysSerLysSerGl 904

QY 1034 GTCCGAGAGCCCTTGTCCGAC-----ACTGGCTCTGAGGATGTTGGTACTGGAGTAA 1087

Db 904 uArgGluCysProHisSerLysLysArgThrLeuLysGluAsnLeuSerAspHisLeuAr 924

QY 1088 AAATGCCCAACAGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCCTCCATTTGAGAA 1147

Db 924 gAsnGlySerLysProLysArgLys-----AsnTyrAlaGlySerLysTr 939

QY 1148 AGAAAGTGAACCTGAGTCACCAATGGATGTA---GATAATTCAAAAATAGTTGTACGGA 1204

Db 939 pAspSerGluSerAsnSerGluArgAspValThrLysAsnSerLysAsnAspSerHisPr 959

QY 1205 TTCAGAACGAGATGAAGAGACAAAGTCCAGGTTTGTATGAACAGAGAAATAGCAGTTCTGC 1264

Db 959 oSerSerAspLysGluGluGlyAlaThrSerAspSerGluSerGluValSerGluIl 979

QY 1265 TCAACAGCAATAAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTTGAGGAA 1324

Db 979 eHisIleLysValLysProThr-----Th 987

QY 1325 GCGTCTCTGCTAAGGGAGGTGAGATTGATTACATTTCCATTGAAGGAGGAGAGAG 1384

Db 987 rLysSerSer-----Th 991

QY 1385 TCGAGCTGGATGAATGATGTGAAT-----GCCAAACGACCTGGAAGTAC 1429  
Db 991 rAsnThrSerLeuProAspAspAsnGlyAlaTrpLysSerSerLysGlnArgThrSerTh 1011  
QY 1430 TTCTAGCCTGAATGTAGAGTGCAGAAATCTTAAG----- 1463  
Db 1011 rSerAspSerGluGlySerCysSerAsnSerGluAsnAsnArgGlyLysProGlnLysHi 1031  
QY 1464 -CAACATGGGAGAAAGGATTCTAAATATCAGATCATTTTCATGAGAGTGCCCAAAGCAGA 1522  
Db 1031 sLysHisGlySerLysGluAsnLeuLysArgGluHis-----ThrLysLysVally 1048  
QY 1523 GGACAAAAGAAAAGAACAAATGTGAAATGAACATCAAGAACAGAAAGGAAGATCCCTAA 1582  
Db 1048 sGluLysLeuLysGlyLysLysAspLysLysHisLysAlaProLysArgLysGlnAlaPh 1068  
QY 1583 ATACATTCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACCTCTATTGAGGAGAT 1642  
Db 1068 eHisTrpGlnPro----- 1072  
QY 1643 GAGGAGAAATGCCAAGGTGTGGATCCGGCTGCCCTTGAGACCATCTGCCAATCACAC 1702  
Db 1072 ----- 1072  
QY 1703 AGTGACTATTTCGGGTAGATCTTTTTCGAATAGGAGAAGTTCTTAAACCTTTCCCAACACA 1762  
Db 1073 -----ProLeuGluPheGlyGluGl 1079  
QY 1763 TTTTAAAGATTTGTGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAACCTTGTA 1822  
Db 1079 uGluGluGluGluLeuAspAspLysGlnValThrGlnGluSerLysGluLysLysVal-- 1098  
QY 1823 CCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACATTCAGAC 1882  
Db 1099 -----SerGluAsnAsnGlu----- 1103  
QY 1883 TGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTTGAAGTACAA 1942  
Db 1104 -----ThrileLysAspAsnilleLysThrGl 1113  
QY 1943 TGTGGCA-----TATTCTAAGAAATGGGACTTTTACAGCTTTTGATTGATTTCTG 1990  
Db 1113 uLysSerSerGluGluAspLeuSerGlyLysHisAspThrValThrValSerSerAspLe 1133  
QY 1991 GGATAAGGTACTAGAGAAGCAGAAAGCTCAACACTGTATCATGTCCTCATCTTGCCTGATAT 2050  
Db 1133 uAspGlnPheThrLysAspAspSerLysLeuSerLysSerProThrAlaLeuAsnThrGl 1153  
QY 2051 GGTGAAAATTCGACTCTGTCTGCCAAATATTGTGACCCAGCCCAATACCACTCCTGAAACA 2110  
Db 1153 uGluAsnValAla---CysLeuGlnAsnIle-----G1 1163  
QY 2111 GAAGATGAATCATTCATCACAATGTCCAGGAACAGATGTCAGTCTTTTAGCTAATGC 2170  
Db 1163 nHisValGluGluSerValProAsnGlyValGluAspValLeuGlnThrAspAspAsnMe 1183  
QY 2171 TTTCTTCTGCAGCTTTCAGCAGCAATGCCAAGATGAAATCAGAGTATTCAGT----- 2225  
Db 1183 tGluileCysThr---ProAspArgSerSerProAlaLysValGluGluThrSerProLe 1202  
QY 2226 -----TATCCAGATATTAAAC 2240  
Db 1202 uGlyAsnAlaArgLeuAspThrProAspIleAsn 1213

RESULT 10  
I39410  
AF-4 protein, splice form 3 - human  
N;Alternate names: serine/proline-rich FEL protein  
C;Species: Homo sapiens (man)  
C;Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 29-Aug-1997  
C;Accession: I39410  
R;Nakamura, T.; Alder, H.; Gu, Y.; Prasad, R.; Canaani, O.; Kamada, N.; Gale, R.P.; Lang  
Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993

A;Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leu  
A;Reference number: A47440; MUID:93281633; PMID:8506309  
A;Accession: I39410  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1210 <NAK>  
A;Cross-references: GB:L13773; NID:G306446; PID:G306447  
C;Comment: This protein is one of several to form chimeric fusion proteins with the trit  
C;Genetics:  
A;Gene: GDB:MLLT2; AF-4; FEL  
A;Cross-references: GDB:136792; OMIM:159557  
A;Map position: 4q21-4q21  
C;Keywords: alternative splicing  
Alignment Scores:  
Pred. No.: 0.000875 Length: 1210  
Score: 173.00 Matches: 227  
Percent Similarity: 32.79% Conservative: 137  
Best Local Similarity: 20.45% Mismatches: 453  
Query Match: 2.39% Indels: 293  
DB: 2 Gaps: 53

US-09-302-812-1 (1-4070) x I39410 (1-1210)

QY 79 CGTCTGATAGGCGCTGGTTTCGGGAGGCTGTTCAGAGCAGGAGTCTCAGAAAGCAGTCAGCG 138  
Db 163 ArgLeuGlyGlnGluGlyPheGly-----SerSerHisHisLysLysGlyAspArgArg 180  
QY 139 GCAGAGGGGCGATGGTCCCG-----GAGGCACCGAGGAGGGGGCGCAGTCCG 186  
Db 181 AlaAspGlyAspHisCysAlaSerValThrAspSerAlaProGluArgGluLeuSerPro 200  
QY 187 -----TCCTCCAGGGTTAGTG-----AATGAGGCTCTA 216  
Db 201 LeuIleSerLeuProSerProValProLeuSerProIleHisSerAsnGlnGlnThr 220  
QY 217 CGCCCGGCTGGCCCGGAGACTCAGTGTGGGTCCCGTCCAGCATGAGTCCGG----- 267  
Db 221 LeuProArgThrGlnGlySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyr 240  
QY 268 GCCCGGCTGTGAGCCCT-----GCACCAAGC 294  
Db 241 CysProAlaLysSerProLysAspLeuAlaValLysValHisAspLysGluThrProGln 260  
QY 295 GAC-----CCCGTGGGAGCGCGCTGCAACTTCTCCCGCGCGCGCTCGGACG 342  
Db 261 AspSerLeuValAlaProAlaGlnProProSerGlnThrPhe---ProProProSerLeu 279  
QY 343 CCCGAGCTTCCCGGCGCAGCAGAGCGCGCTCCTCGATTCCAGGACGC-----TCCGG 396  
Db 280 ProSerLysSerValAla-MetGlnGlnLysProThrAlaTyrValArgProMetAspGl 299  
QY 397 TGCAGTTTCAGGTCCTCCCGCTCTCTCTGTCAGGCTGCGCTGCGCGCGCGCGG----- 448  
Db 299 yGlnAspGlnAlaProSerGluSerProGluLeuLysProLeuProGluAspTyrArgGl 319  
QY 449 ----ACAGCACCGAGGCGCGCCACCTCTCTGTTTTCAAACAGAGACTATAACAGTT 504  
Db 319 nGlnThrPheGluLys-ThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuL 339  
QY 505 GGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGATAAGAA----- 560  
Db 339 ysMetProSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGl 359  
QY 561 -----AACACAATA 570  
Db 359 leLeuLysGluMetThrHisSerTrpProProProLeuThrAlaIleHisThrProSerT 379  
QY 571 CAAGAGAAGAATCCATGATGATGTTCTGTACAAAAGATAACTTTTATCAACATAACATGG 630  
Db 379 hrAlaGluProSerLysPheProPheProThrLysAspSer-----GlnHisValSerS 397  
QY 631 AAAAATTAGAAATGTTTCTCAGCTAGGTTTGTGATAAGTCCACCAGTTGAAAAAGGTACAC 690



Db 397 erValThrGlnAnGlnLysGln-----TyAspThrSer-----SerLysThrH 412  
QY 691 AGTATTGAAGCAGCATCAGACTCGCGCTATGTGTAAG-----TGGCAGA 735  
Db 412 isSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspLeuGlnLeuSerAsps 432  
QY 736 ATGAAGGCCACACTCAGAACGGCTTTTGGAA-----AGTGAACCTCCA- 779  
Db 432 erGluAspSerAspSerGluGlnThrProGluLysProProSerSerSerAlaProProS 452  
QY 780 --GCGGTAACCTGTGTACAGAGCAGTTTCAGTAATGCTAAT----- 818  
Db 452 erAlaProGlnSerLeuProGluProValAlaSerAlaHisSerSerSerAlaGluSerG 472  
QY 819 -----GTCGATCAGTCGTCCCAAGGATGATCAGTGCACACAATAGTGAGGAGA 870  
Db 472 luSerThrSerAspSerAspSerSerSerAspSerGluSerGluSerSerSerAsps 492  
QY 871 GTAGAGATAATCAGCAGTTTGTGACACATGTAAGCTTGCAATGCAAGCAGCAGATGG 930  
Db 492 erGluGluAsnGlu-----ProLeuGluThrProAlaProGluProG 506  
QY 931 AAGATGAACAGGGCAGAGAAGCCAGAGCCAGGAAGTGTGGCAAGGCTTGCCATCCTG 990  
Db 506 luProProThrThrAsnLysTrpGlnLeuAspAsnTrpLeuThrLysValSerGlnProA 526  
QY 991 CAGAAGCCTGTGCGGTGTGTCAGCAGGAGGAGCAGACGTGGTGTCCGAGAGCCCTTGT 1050  
Db 526 laAlaProProGluGly-----ProArgSerThrGluProProArgArgHisProGluS 544  
QY 1051 CGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAACAGATTGAATAGAC 1110  
Db 544 er---LysGlySerSerAspSerAlaThrSer-----G 554  
QY 1111 AAGAAAGTAGCTAGGAAATCTCTCCATTTGAGAAAGAAAGTGAA---CCTGAGTCAC 1167  
Db 554 lnGluHisSerGluSerLysAspProProProLysSerSerSerLysAlaProArgAlaP 574  
QY 1168 CAATGGATGTAGATAAT---TCCAAAAATAGTTGTGAGGATTGAGAAAGCAGATGAAGAGA 1224  
Db 574 roProGluAlaProHisProGlyLysArgSerCysGlnLysSerProAla----- 590  
QY 1225 CAAGTCCAGGTTTGTGATGAACAGGAAGATAGCAGTTCTGCTCAAAACAGCAATAAACCTT 1284  
Db 591 -----GlnGlnGluProProGlnArgGlnThrValGlyThrLysG 604  
QY 1285 CAAGGTTCCAAACAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCTCTGTGAAGGAG 1344  
Db 604 lnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSerLeuGlnGly- 623  
QY 1345 GTGAGATTGATTTCAATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATGAATGATG 1404  
Db 624 -----GluArgGluProGlyLeuLeuProTyArgSerArgAspG 637  
QY 1405 TGAATGCCAAA---CGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATCTTA 1461  
Db 637 lnThrSerLysAspLysPro-----LysValL 646  
QY 1462 AGCAACATGGGAGA---AAGGATTCTAAATACAGATCATTTTCATGAGAGTGCCCAAAG 1518  
Db 646 ysThrLysGlyArgProArgAlaAlaAlaSerAsnGluProLysProAlaValProProS 666  
QY 1519 CAGAGGACAAAAGAAAAGACAATGT----- 1544  
Db 666 erSerGluLysLysLysHisLysSerSerLeuProAlaProSerLysAlaLeuSerGlyP 686  
QY 1545 --GAAATGAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACATT----- 1589  
Db 686 roGluProAlaLysAspAsnValGluAspArgThrProGluHisPheAlaLeuValProL 706  
QY 1590 -----CCACCTCACCTTTTCTCCAGATAAGAAATGGCTTGA----- 1625

Db 706 euThrGluSerGlnGlyProProHisSerGlySerGlySerArgThrSerGlyCysArgG 726  
QY 1626 -----ACTCCTATTGAGGAGATGAGGAGAATGCCAAGGTGTGGGATC----- 1667  
Db 726 lnAlaValValGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspT 746  
QY 1668 -----CGGCTGCCTCCCTTGAGA-----CCATCTGCCAATCAGCAGTGACTATTTC 1713  
Db 746 hrLysLeuLeuSerProLeuArgAspThrProProProGlnSerLeuMetValLysIleT 766  
QY 1714 GGGTAGATCTTTTTCGAATAGGAGAAGTTCTTAAACCTTTTCCCAACACATTTTAAAGATT 1773  
Db 766 hrLeuAspLeuLeu-----SerArgIleProGlnPro----- 776  
QY 1774 TGTGGGACAAACAAGCATGTTAAGATGCCTTGTTCAGAAACAAAACCTTGACCTGTG---G 1830  
Db 777 -----ProGlyLysGlySerArgGlnArgLysAlaGluAspLysGlnProProAlaGlyL 795  
QY 1831 AAGATGAGATGGTGTGAGCGAGCTGCAGGAGCCGGTGGGAACCTCAATCAGACTGCACCTTC 1890  
Db 795 ysLysHisSerSerGluLysArgSerSerSerAspSerSerSerLysLeuAlaLysLysArgL 815  
QY 1891 TCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAGTACAATGTGGCAT 1950  
Db 815 ysGlyGluAlaGluArgAspCysAspAsnLysLysIleArgLeuGluLysGluIleLysS 835  
QY 1951 ATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTCTGGGATAAGGTACTAGAAGAAG 2010  
Db 835 erGlnSerSerSerSerSerSerHisLysGluSerSerLysThrLysProSerArgP 855  
QY 2011 CAGAAGCTCAACACTTGTATCAGTCCATCTTCCTGCTGATATGGTGAAAATTCGCACTCTGTC 2070  
Db 855 roSerSerGlnSerSerLysLysGluMetLeuProPro----- 867  
QY 2071 TGCCAAATATTGTACC-----CAGCCAATA---CCACTCCTGAAACAGAAAGATGA 2118  
Db 868 --ProProValSerSerSerSerGlnLysProAlaLysProAlaLeuLysArgSerArgA 887  
QY 2119 ATCATTCATCCATCAATGTCACAGGAA-----CAGATTGCCAGTCTTTTAGCTAATG 2169  
Db 887 rgGluAlaAspThrCysGlyGlnAspProProLysSerAlaSerSerThrLysSerAsnH 907  
QY 2170 CTTTCTTCTGCACTTTCCACGA-----CGCAATGCCAAGATGAAAT 2211  
Db 907 isLysAspSerSerIleProLysGlnArgValGluGlyLysGlySerArgSerS 927  
QY 2212 CAGAGTATTCC----- 2222  
Db 927 erGluHisLysGlySerSerGlyAspThrAlaAsnProPheProValProSerLeuProA 947  
QY 2223 -----AGTTATCCAGATATTAACTTCAATCGGTTGTTGAAGGACGTT 2265  
Db 947 snGlyAsnSerLysProGlyLysProGlnValLysPheAspLysGln-GlnAlaAspLeu 966  
QY 2266 CATCAAGGAAACCAGAGAAGCTTAAACCGCTCTTCTGCTACTTTTAGAAGAGTCACAGAGA 2325  
Db 967 HisMetArgGluAlaLysLysMetLysGlnLysAlaGluLeuMet----- 981  
QY 2326 AAAAACCCACTGGTGGTGACATTTCACAAGACAGAGTCTTTGAAGATTTCACAGAGTGGG 2385  
Db 982 -----ThrAspArgValGlyLysAlaPheLysTyrLeu 992  
QY 2386 AAAGATGTGAAAACTCTGACTCGACTGCATGCTCACTTACGAAGGTACCATAGAAGGAA 2445  
Db 993 GluAlaValLeu---SerPheIleGluCysGlyIleAlaThrGluSerGlu---SerGln 1010  
QY 2446 ACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAACCGTTTCGTTG----- 2490  
Db 1011 SerSerLysSerAlaTyrSerValTyrSerGluThrValAspLeuIleLysPheIleMet 1030  
QY 2491 -----GAGGTGGTGTAAACCAAGTGCAGGAC----- 2514  
Db 1031 SerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlaValLeu 1050



QY 2515 -----TTGTGCAAGAAGAAATCCGCT 2535  
Db 1051 CysMetArgCysGlnSerIleLeuAsnMetAlaMetPheArgCysLysLysAspIleAla 1070  
QY 2536 -----TTTTAATCAACCCCTGAGTTGATTTTTCACGGCTCTTCACTGAGGTGCTGGATC 2589  
Db 1071 IleLysTyrSerArgThrLeuAsnLysHisPheGluSerSerLysValAlaGlnAla 1090  
QY 2590 ACAATGAATGCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCG 2649  
Db 1091 ProSerProCysIleAlaSerThrGlyThrProSerProLeuSerProMetProSerPro 1110  
QY 2650 AACATACCGCTGGGCCCGGAGCCATGAGACAGGAGCGAAGGGACGACTGGCAGAGGC 2709  
Db 1111 AlaSerSerValGly---SerGlnSerSerAlaGlySerValGlySerSerGlyValAla 1129  
QY 2710 GCACGACTGAGATCGTGCCTCATCGACGCCCTCCACT-----TCAGAC 2751  
Db 1130 AlaThrIle-----SerThrProValThrIleGlnAsnMetThrSerSerTyr 1145  
QY 2752 GCTACCTCGACCACTTTGTGCCCCGAGAGATCAGAC-----GGGAGCTTA 2796  
Db 1146 ValThrIleThrSerHisValLeuThrAlaPheAspLeuTrpGluGlnAlaGluAlaLeu 1165  
QY 2797 ACAAGGCTTACTGTGGATTCTTC 2820  
Db 1166 ThrArgLysAsnLysGluPhePhe 1173  
RESULT 11  
A46194  
Neurofilament protein NF-220, high-molecular-weight splice form - longfin squid  
C;Species: Loligo pealeii (longfin squid)  
C;Date: 22-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 17-Mar-2000  
C;Accession: A46194  
R;Way, J.; Hellmich, M.R.; Jaffe, H.; Szaro, B.; Pant, H.C.; Gainer, H.; Battey, J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6963-6967, 1992  
A;Title: A high-molecular-weight squid neurofilament protein contains a lamin-like rod  
A;Reference number: A46194; MUID:92357751; PMID:1379729  
A;Accession: A46194  
A;Status: nucleic acid sequence not shown; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-1200 <WAY>  
A;Cross-references: GB:M94389; NID:g161291; PID:g161292  
A;Experimental source: stellate ganglion  
A;Note: sequence extracted from NCBI backbone (NCBIP:113499)  
C;Superfamily: neurofilament triplet H protein  
C;Keywords: alternative splicing  
Alignment Scores:  
Pred. No.: 0.0013 Length: 1200  
Score: 170.50 Matches: 116  
Percent Similarity: 36.35% Conservative: 85  
Best Local Similarity: 20.98% Mismatches: 257  
Query Match: 2.35% Indels: 95  
DB: 2 Gaps: 21  
US-09-302-812-1 (1-4070) x A46194 (1-1200)  
QY 178 CGCAGTCCGTCCTCCAGGTTAGTGAATGAGGCTTACGCCCGGGTGGCCCGGAGAC 237  
Db 644 LysSerProLysSerProGlySerLysArgAspThrLeu-----PheGlnSerAsp 660  
QY 238 TCAGTGTGCGGTCCAGCATGAGTGGCGGCCCGGCTGTG----- 279  
Db 661 LysValThrSerProValLeuLeuGluAlaMetSerLeuGlnCysAlaProThrGlu 680  
QY 280 -----AGCCCTGCACCAAGGACCCCGCTGGGACGCCGCTG 315  
Db 681 LysProLeuSerAspThrIleLysSerProValValSerGluProAlaIleSerProVal 700  
QY 316 CAACTTCTCCGCCCGCTCGGACGCCCGGAGCT-----TCCCGGCGAGGC 363

Db 701 SerSerValArgSerGlyAlaGluThrSerProAlaGluArgMetThrSerProIleGly 720  
QY 364 AGAGGCGCGTCTCGATTCCAAGGACGCTCCGGTGCAGTTACGGTCCCGCTCCTCGT 423  
Db 721 SerGlySerGluLysSerAlaLysSerProValArgSerGluAlaThrLysSerPro--- 739  
QY 424 CAGGTCGCGCCCTGGCGCGGACAGCACCAGGAGCGCGCACCTCTCTTGTTCATCA 483  
Db 740 ValSerGluLysSerAlaGlySerLysSerLysSerProValPro-----Ser 755  
QY 484 AACAGAGACTATAACCAAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAA 543  
Db 756 GluArgAlaLeuSerProVal-----LeuSerGlu--MetSerValHisSerThrA 772  
QY 544 GTTTG---CATAGTAAAGAAAACAACAATACAGAGAAGATCCCATGATGATTCTGTAC 600  
Db 772 laMetSerHisThrSerArgSerProThrAlaSerGluLysSerValLysSerProThrH 792  
QY 601 AAAAGATAACTTTTATCAACATAACATGGAATA-----TTAGAAAATGTTTCTC 651  
Db 792 isSerGluArgThrAlaSerProThrAlaLysSerProIleIleMetGluProAla--- 810  
QY 652 AGCTAGGTTTGTATAGTACACAGTTGAAAGAGGTACACAGTATTTTGAAGCAGCATCAGA 711  
Db 811 -----LysSerProLysAspGluSerGluLysGluLeuSerProGluArgS 826  
QY 712 CTGCGGCTATGTAAAGTGGCAGAAATGAA-----GGGCCACACT 750  
Db 826 erGluSerValMetSerGlyLysSerGlnIleThrSerSerSerAlaLysSerProValP 846  
QY 751 CAGAACCGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCAGACAGTTCAGTA 810  
Db 846 roSerGluLysAlaAspSerGluLysSerAlaThrSerProThrProSerGlu-----L 864  
QY 811 ATGCTAATGTCGATCAGTCGTGCTCCCAAGGATGATCAGGTGAC---ACAAATAGTGAGG 867  
Db 864 ysValAspAspGluSerSerAlaArgSerProThrHisSerAspArgSerGluSerGlnL 884  
QY 868 AGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAGCTTGCAGATGCAAGCAGACGA 927  
Db 884 ysSerAlaArgSerProMetThrSerSerHisIleLysSerProIleAspPheGluLysA 904  
QY 928 TGAAGATGAACAGGGC----- 944  
Db 904 laGluSerGluLysSerAlaLeuSerTyLysSerSerAspHisAspLysSerProValP 924  
QY 945 --AGAGAAGCCAGAACCCACCAGAGTGTGGCAAGGTTGCCATCCTCGAGAAGCCTGTG 1002  
Db 924 roSerGluLysAlaGluSerGluLysSerAlaArgSerProValProSerGluLysAlaG 944  
QY 1003 CAGGGTGTCCAGCAGGAGGAGACAGACGTCGTGTCCTCCGAGAGCCCCCTTGTCCGACACTGGCT 1062  
Db 944 luSerGluLysSerAlaArgSerProValThrSerGluProAlaLysSerProValProS 964  
QY 1063 CTGAGGATGTTGTGCTGACTGAAATAATGCC-----AACAGATTGAATA 1107  
Db 964 erGluLysAlaAlaGluSerGluLysSerAlaArgSerProValProSerGluLysAlaG 984  
QY 1108 GACAAGAAAGTAGTCTAGGAAATTTCTCTCTCCATTTGAGAAA---GAAAGTGAACCTGAGT 1164  
Db 984 luSerGluLysSerAlaArgSerProValProSerGluLysAlaGluSerGluLysSera 1004  
QY 1165 CA-----CCAAATGGATGTAGATAATTCCTCTCCATTTGAGAAAATAGTTGTCCAGGATTCAGAAG 1212  
Db 1004 laArgSerProValProSerGluLysAspGluSerGluLysSerAlaLysSerProThrS 1024  
QY 1213 CAGATGAAGAGACAAAGTCCAGGTTTGTATGAACAGGAAGATAGCAGTTCTGTCTCAACAG 1272  
Db 1024 erAspGlnAlaArgSerPro-----ValProSerGluLysAlaGluSerGluLysSera 1042  
QY 1273 CAAATAAACCT-----TCAAGGTTCCAACCAAGAGAGAGCTGACA 1311  
Db 1042 laArgSerProValThrSerGluHisValLysSerProValProSerGluLysSerGluS 1062

QY 1312 CTGAGTTGAGGAGCGGTCTCTGCTAAGGAGGTGAGATTGATATACATTTCCAATTG 1371  
Db 1062 erGluLysSerAlaArgSerProValThrSerGluHisValLysSerProValProSerG 1082  
QY 1372 AAGGAGGAGAGTCGAGCTGGAATGAATGAATGAATGGAATGCCAAACGACCTGGAAGTACTT 1431  
Db 1082 luLysAlaGluSer-----GluLysSerAlaLysSerProValAlaSerG 1097  
QY 1432 CTAGCCTGAATGTAGAGTCAGAAATTTCTAAGCAACATGGGAGAAAGGATTTCTAAA---- 1487  
Db 1097 luProAlaLysSerProIleProSerGluLysAlaGluSerGluLysSerAlaLysSerP 1117  
QY 1488 --ATCACAGATCATTTTCATGAGAGTGCCCAAGCAGAGGACAAAAGA-----AAAG 1536  
Db 1117 roValThrSerGluHisAlaThrSerProValAlaSerGluLysSerGluSerGlyLysS 1137  
QY 1537 AACAAATGTGAATGAACATCAAGAACAGAGGAAGATCCCTAAATACATTTCCACCTC 1596  
Db 1137 erLysSerProAlaAlaSerGluLysAlaGluSerGluLysSerLysSerProValProS 1157  
QY 1597 ACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCT 1631  
Db 1157 erGluLysGluGlySerLysThrAlaGlySerPro 1168  
  
RESULT 12  
B47328  
natural killer cell tumor-recognition protein - mouse  
N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tum  
C;Species: Mus musculus (house mouse)  
C;Date: 17-Jul-1994 #sequence\_revision 17-Jul-1994 #text\_change 05-Nov-1999  
R;Accession: B47328; I77662  
R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.  
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993  
A;Title: A cyclophilin-related protein involved in the function of natural killer cells.  
A;Reference number: A47328; MUID:93133824; PMID:8421688  
A;Accession: B47328  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1507 <AND>  
A;Cross-references: GB:L04289; NID:g192866  
A;Note: authors translated the codon AGT for residue 972 as Arg  
R;Rinfret, A.; Anderson, S.K.  
Mol. Immunol. 30, 1307-1313, 1993  
A;Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing m  
A;Reference number: I57820; MUID:94019422; PMID:8413330  
A;Accession: I77662  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 235-237;263-294 <RIN>  
A;Cross-references: GB:S65998; NID:g425701; PIDN:AAB28500.1; PID:g425702  
C;Genetics:  
A;Gene: NK-TR  
C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology  
C;Keywords: alternative splicing; lymphocyte  
F;60-230/Domain: cyclophilin homology <CYP>  
  
Alignment Scores:  
Pred. No.: 0.00137 Length: 1507  
Score: 170.50 Matches: 146  
Percent Similarity: 33.38% Conservative: 91  
Best Local Similarity: 20.56% Mismatches: 301  
Query Match: 2.35% Indels: 172  
DB: 2 Gaps: 31  
  
US-09-302-812-1 (1-4070) x B47328 (1-1507)  
QY 101 GGAGGCTGTCAGAGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGCATGGTGCCGGG 160  
Db 562 GlySerAlaSerThrHisSerSerArgAspSerTyrArgSerLysSerHisSerArgSer 581  
QY 161 AGGCACGAGGAGGGGGC-----GCAGTCCGTCCCTCCAGGGTTAGTGAATG--- 208  
:::  
|||||

Db 582 AspSerArgGlySerSerArgSerArgAlaValSerLysSerSerSerArgSerLeuAsn 601  
QY 209 -----AGGCTCTACGCCCGGGCTGGCCCGGAGACTCAGTGTGCGGGTCCCAG 256  
Db 602 ArgSerLysSerArgSerSerArgSerGlyPro-ArgArgThrSerIleSerProLy 621  
QY 257 CATGAGTGGGGCCCCGGC-----TGTGAGCCTCGCACCAAGCGACC 298  
Db 621 sLysProAlaGlnLeuSerGluAsnLysProValLysThrGluProLeuArgProSerVa 641  
QY 299 CCGCTGGGACGCCGCTGCAACTTCTCCCGCGCGCTCGGACGCCCGGAGCTTCCCQGG 358  
Db 641 lProGlnAsnGlyAsnValLeuValGlnProValAla-----AlaGluAsnIlePro-- 658  
QY 359 CAGGCAGAGCGCGCTCCTCGATTCCCAAGGACGCTCCGGTGCAGTTCAGGCTCCCGCGTC 418  
Db 659 -----ValIleProLeuSerAspSerProProSerArgTrpLysProGl 674  
QY 419 CTCGTACGGCTGC-----GCCCTGGCGCGCGGACAGCACCCGAGGCGCCACCTC 472  
Db 674 yGlnLysProTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrThrHi 694  
QY 473 TCTTGTGTTTCAAAACAGAAAGACTATAACCACTGGATGGACACTAAAGGAATCAAG---AC 529  
Db 694 sLeuLeuProValGlnSerThrTyrSer-----LeuThrAsnIleLysAlaTh 710  
QY 530 AGTTGAATCAGAAAGTTTGCATAGTAAAGAAACAAACAATACAAGAGAAATCCATGAT 589  
Db 710 rValSerSerSerTyrHisLysArgGluLysProSerGluSerAspGlySerAlaTy 730  
QY 590 GAGTTCTGTACAAAAAGATAACTTTTATCAACATAAATGAAAAAATTAGAAAAATGTTTC 649  
Db 730 rSerLysTyrSer----- 734  
QY 702 ----CAGCATCAGACTCGGCTATGTGTAGTGGCAGAAATGAAGGGCCACACTCAGAACG 757  
Db 750 gSerArgSerSerArgSerTyrThrArgSerArgSerArgSerLeuProThrSerAr 770  
QY 758 GCTTTTGGAAAGTGAACCTCCAGCGGTAACTTGGTACCAGAGCAGTTTCAATGCTAA 817  
Db 770 gSerLeuSerArgSerProSerSerArgSerHisSerProAsnLysTyrSerAspGlySe 790  
QY 818 TGTCGATCAGTCTCCCAAGGATGATCACAGTCACACAAATAGTGAGGAGACTAGA-- 875  
Db 790 r-----GlnHisSerArgSerSerTyrThrSerValSerSerAspGlyArgAr 808  
QY 876 -----GATAATCAGCAGTTTTTTCACATGTAAAG-----CTTCCGAA 913  
Db 808 gAlaMetPheArgSerAsnArgLysLysSerValThrSerHisLysArgHisArgSerAs 828  
QY 914 TGCAAGCAGACGATGGAAGATGAA-----CAGGCAGAGAGAGCCAGAACG---CACC 964  
Db 828 nSerGluLysThrLeuHisSerLysTyrValArgGlyArgGluLysSerSerArgHisAr 848  
QY 965 GAAAGTGGCAAGGCTTGCCATCTCTGCAGAAAGCCTGTGCAGGGTGTGCAGGAGGAGAC 1024  
Db 848 gLysTyrSerGluSerArgSerSerLeuAspTyrThrSerAspSerAspGlnSerHisVa 868  
QY 1025 AGACGTGGTGTCCGAGAGCCCTTGTGGACACCTGGCTCTGAGGATGTTGGTACTGACT 1084  
Db 868 lGlnValTyrSerAlaProGluLysGluLysGlnGlyValGluAlaLeuAsnAspLy 888  
QY 1085 GAAAAATGCCAACAGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCCTCCTCATTTGA 1144  
Db 888 sGlnGlyLysGlyArgGluGluGlyLysProLysProGluTrpGluCysProArgSerLy 908  
QY 1145 GAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGATAAATCCAAA---AATAGTTGTCA 1201  
Db 908 sLysGluAsnSerGluAspHisSerArgAspSerValSerLysGlyLysAsnCysAl 928  
:::  
|||||





Db 901 uAspGluPheGluThrTyrAsnValAspGluLeuGluAsnValIleAsnAspAspIle 921  
:::|  
QY 1031 GGTGTCGAGAGCCCTTGTCTGGACACTGGCTCTGAG--GATGTTGGTACTGGACTGAA 1087  
:|::|  
Db 921 eAlaGluAlaAsnProLeuProAspGluAsnAspValGlnMetAsnGluSerPheAs 941  
:::|  
QY 1088 AAATGCCAACAGATTGAATAGACAGAAAGTAGTCTAGGAAATTCTCTCCATTGAGAA 1147  
|||:::|  
Db 941 pAsnAsnHisSerMetSerArgAlaLysLysLysTyr-----ThrPheGluLy 957  
|||:::|  
QY 1148 AGAAAGTGAACTGAGTCACCAATGGATGTAGATAATTCCAAAATAGTTGTTCAGGATTC 1207  
|||:::|  
Db 957 sGluValAsnGluLys-----IleAlaGlyThrLysHisSerLeu----- 970  
|||:::|  
QY 1208 AGAAGCAGATGAAGACAGACAAAGTCCAGGTTTGTGATGAACAGGAGATAGCAGTTCTGCTCA 1267  
:::|  
Db 971 -----AspThrThrAspPro-----ArgGluAlaIleArgValLe 982  
:::|  
QY 1268 AACAGCAATAAACCTTCAAGGTTCCAAACCAAGAGAGAGCTGAC-----ACTGA 1315  
:::|  
Db 982 uAsnThrGlyGluThrLysArgIleGluProLysLysArgGluValProIleThrValLy 1002  
:::|  
QY 1316 GTTGAGGAAGCGGTCCTCTGCTAAGGGAGGTGAGATTGATTACATTTCCTAATTGAAGG 1375  
:::|  
Db 1002 sLeuAsnLysArgSerGlnTyrLysSerProTyrVal----- 1014  
:::|  
QY 1376 AGGAGAGAGTCGAGCTGGAATGAATGATGTGAATGCCAAACGA----- 1418  
:::|  
Db 1015 -----ThrArgSerGlyArgThrValIleAsnProLysArgTyrLeuHisAlaValVa 1032  
:::|  
QY 1419 -----CCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAA 1456  
|||  
Db 1032 lAsnLysIleAspTyrAsnAspProGlyTyrIleLysSerMetAsnAlaGluLeuGluLy 1052  
:::|  
QY 1457 TTCTAAGCAACATGGGAGAAAGGAT----- 1483  
:::|  
Db 1052 sPheArgSer-----LysAspValTyrGluGluValProIleProThrGlyValLy 1069  
:::|  
QY 1484 TAAATACAGATCATTTTCATGAGAGTGCCCAAGCAGAGGAC-----AAAAG 1531  
|||  
Db 1069 sProIleSerMetGlyTyrValHisThrGluLysIleAspSerLeuLysGlyValValAr 1089  
:::|  
QY 1532 AAAAGAACAATGTGAATGAACATCAAGAAGCAAGAAAGGAGATC----- 1577  
|||  
Db 1089 GlySerArgCysValValHisGlyAsnArgGlnLysGluLysLeuAspTyrAspPropH 1109  
:::|  
QY 1577 ----- 1577  
Db 1109 eSerValSerSerProValIleAspLeuValThrIleArgLeuLeuThrIleIleGlyCy 1129  
1577 ----- 1577  
Db 1129 sGluLeuGlyMetThrIleGlnHisLeuAspValGluSerAlaTyrLeuAsnAlaSeril 1149  
1578 -----CCTAAATACATT---CCACCTCACCTTCTCCAGATAAGAAA----- 1616  
|||  
Db 1149 eThrHisSerAsnProIleTyrValPheProProLysSerValProLeuLysLysAsnHi 1169  
|||  
QY 1617 -----TGGCTTGAACCTCTATTGAGGAGATGAGGAGA-----ATGCCAAGGTG 1660  
|||  
Db 1169 sCysTrpLeu-----LeuLysArgSerValTyrGlyLeuLysGlnSe 1183  
|||  
QY 1661 TGGATCCGGTGCCTCCCTTGAGACCATTGCCCAATCACACAGTACTATTTCGGGTA-- 1718  
|||  
Db 1183 rGlyLeuGluTrp-----TyrHisThrIleLys---ArgValle 1195  
|||  
QY 1718 ----- 1718  
Db 1195 uGluAspIleGlyPheThrGlnValLeuHisAsnAspGlyLeuPheHisIleGluTyrGl 1215  
1719 -----GATCTTTTGGCAATAGGAGAAGT 1741  
|||

Db 1215 uGluGlySerValIleTyrLeuGlyLeuTyrValAspIleLeuMetValGlySerSe 1235  
QY 1742 TCCTAAACCTTTCCCAACACATTTTAAAGATTGTGGGACAAACAAGCATGTTAAGATGCC 1801  
|||  
Db 1235 rGlnLysValIleAspAsnPheValAspGlnLeuArgAspHisPheGluValLysValPh 1255  
|||  
QY 1802 TTGTTTCAGAAACAAACTTGTACCTCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAG 1861  
|||  
Db 1255 eGlyGluIleSerAsnTyrLeuGlyIleGluPheArgLysThrGlu-----Se 1271  
|||  
QY 1862 CCGTGGGAACACTCATTTCAGACTGCACCTTCTCAACAGGCTCCTCGSCCCAGAACCTGAA 1921  
|||  
Db 1271 rGlyTyrIleLeuSerGlnGluLysPheLeuLysLysLeuLysAspPheLysLeuAs 1291  
|||  
QY 1922 GGATGCTATTCTGAAGTACAATGTGGCATAT-----TCTAAGAAATGGGACTTTACAGC 1975  
|||  
Db 1291 pAspSerTyrGlyLys---AsnIleProTyrIleProAsnAspLysTyrGluLysValAl 1310  
|||  
QY 1976 TTTGATT-----GATTCTGGGATAAGGTA----- 2000  
|||  
Db 1310 aIleIleArgGluAsnValAsnProGluAsnAspPhe---GluLysValProAsnGluTh 1329  
|||  
QY 2001 -CTAGAAGAAGCAGAAAGCTCAACACTTGTATCAGTCC----- 2036  
|||  
Db 1329 rLeuLeuAspProAspAlaLysLysLysLeuTyrGlnSerGlyValGlySerLeuLeuTrpAl 1349  
|||  
QY 2037 -----ATCTTGCTGATATGGTGAATTTGCACCTCTGTCTGCCAAATATTGTGTAC 2086  
|||  
Db 1349 aAlaThrAsnThrArgProAspIleSerValValValAsnSerLeuGlySerLysSerAl 1369  
|||  
QY 2087 CCAGCCA-----ATACCACCTCCTGAAACAGAA 2113  
|||  
Db 1369 aAsnProAsnValHisAspTyrGluLysLeuIleTyrCysLeuArgTyrIleLysAsnSe 1389  
|||  
QY 2114 GATGAATCATTCATCACAATGTACACAGGAACAGATTGCCAGTC 2157  
|||  
Db 1389 rMetGlyTyrHisIle-GluTyrLysArgAsnArgLeuAsnIle 1403  
|||

RESULT 14

T29757 protein UNC-89 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Dec-1999

C;Accession: T29757

R;Du, Z.; Le, T.T.; Wilson, R.

submitted to the EMBL Data Library, May 1997

A;Description: The sequence of C. elegans cosmid C09D1.

A;Reference number: Z20679

A;Accession: T29757

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-6642 <DUZ>

A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89

A;Experimental source: strain Bristol N2; clone C09D1

C;Genetics:

A;Gene: CESP:unc-89

A;Map position: 1

A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 61

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Alignment Scores:

|                        |         |               |      |
|------------------------|---------|---------------|------|
| Pred. No.:             | 0.00196 | Length:       | 6642 |
| Score:                 | 170.50  | Matches:      | 198  |
| Percent Similarity:    | 35.12%  | Conservative: | 142  |
| Best Local Similarity: | 20.45%  | Mismatches:   | 418  |
| Query Match:           | 2.35%   | Indels:       | 210  |
| DB:                    | 2       | Gaps:         | 47   |

US-09-302-812-1 (1-4070) x T29757 (1-6642)

QY 246 GCGGGTCCAGCATGAGTCGGGCCCCGGCTGTGAG---CCCTGCACCAAGCAGCCCGC 302

Db 1394 AlaserProThrLysLysThrGlyGluValLysSerProLysGluLysSerProAla 1413  
|||



|    |      |   |      |
|----|------|---|------|
| QY | 303  | TGGGACGCCGCTGCAACTTCTCCGCGCGCCCTCGGACGCCCGAGCTTCCCGGCAGG      | 362  |
| Db | 1414 | SerProThrLysLysGluLysSerProAlaAlaGluGluVallysSer---ProThrLys  | 1432 |
| QY | 363  | CAGAGGGCGCTCCTCGATTCCAAGGACGCTCCGGTGCAGTTCAGGTCGCCGCGCTCCTCG  | 422  |
| Db | 1433 | LysGluLys-----SerProSerSerProThrLysLysGluLysSerProSerSer      | 1449 |
| QY | 423  | TCAGGCTGGCCCTCGGCGGGGACAGCACCGGAGCGCCACCTCTCTGTGTTTC          | 482  |
| Db | 1450 | ProThrLysLysThrGlyAspGluVallysGluLysSerProProLysSerProThrLys  | 1469 |
| QY | 483  | AAACAGAGACTATAACAGTTGGATGGACACTAAA-----                       | 518  |
| Db | 1470 | LysGluLysSerProGluLysProGluAspVallysSerProVallysLysGluLysSer  | 1489 |
| QY | 519  | -----GGATCAAGACAGTTGAATCAGAA---AGTTTGCATAGTAAGAAAAAC          | 563  |
| Db | 1490 | ProAspAlaThrAsnIleValGluValSerSerGluThrThrIleGluLysThrGluThr  | 1509 |
| QY | 564  | AACAATACAAGAGAA-----GAATCCATGATGAGTTCTCTACAAAAAGAT            | 608  |
| Db | 1510 | ThrMetThrThrGluMetThrHisGluSerGluGluSerArgThrSerVallysLysGlu  | 1529 |
| QY | 609  | AACTTTTATCAACATAACATGGAATAATTAGAAATGTTTCTCAGCTAGCTTTTGATAAG   | 668  |
| Db | 1530 | LysThrProGluLysValAspGluLysProLysSerProThrLysLys-----AspLys   | 1547 |
| QY | 669  | TCACCAGTTGAAAAAGGTACACAGTATTGGAAGCAGCATCAGACTGCGGCTATGTGTAAG  | 728  |
| Db | 1548 | SerProGluLysSerIleThrGluGluIleLysSerProVallys-----            | 1562 |
| QY | 729  | TGGCAGAATGAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACT  | 788  |
| Db | 1563 | ---LysGluLysSerProGluLysValGluLysProAlaSerProThrLysLysGlu     | 1581 |
| QY | 789  | CTGCTACCAGACGAGTTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCAC | 848  |
| Db | 1582 | LysSerProGluLys-----ProAlaSerProThrLysLysSerGluAsnGluVallys   | 1599 |
| QY | 849  | AGTGACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAGCTT   | 908  |
| Db | 1600 | SerProThrLysLysGluLysSerProGlu---LysSerValValGluGluLeuLysSer  | 1618 |
| QY | 909  | GCGAATGCAAGCAGACGATGGAA---GATGAACAGGGCAGAGAGCCAGAGCCACCAG     | 965  |
| Db | 1619 | ProLysGluLysSerProGluLysAlaAspLysProLysSerProThrLysLysGlu     | 1638 |
| QY | 966  | AAGTGTGGCAAGCTTGCCATCCTGCAGAAAGCTGTGCAGGGTGTGCAGCAGGAGACACA   | 1025 |
| Db | 1639 | LysSerProGluLysSerAlaThrGluAspVallysSerProThrLysLysGluLysSer  | 1658 |
| QY | 1026 | ---GACGTGTGTCGAGAGCCCTTGTTCGGACACTGGCTCTGAGGATGTTGGTACT---    | 1079 |
| Db | 1659 | ProGluLysValGluLysProThrSerProThrLysLysGluSerSerProThrLys     | 1678 |
| QY | 1080 | -----GGACTGAAAAATGCCAACAGATTGAATAGACAAGAAAGTAGTCTAGGA         | 1127 |
| Db | 1679 | LysThrAspAspGluVallysSerProThrLysLysGluLysSerProGlnThrValGlu  | 1698 |
| QY | 1128 | AATTCTCCT-----CCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGAT   | 1181 |
| Db | 1699 | GluLysProAlaSerProThrLysLysGluLysSerProGlu-----               | 1712 |
| QY | 1182 | AATTCCAAATAATAGTTGTTCAGGATTCAGAAAGCAGATGAAGAGACAAGTCCAGTTTGAT | 1241 |
| Db | 1713 | -----LysSerValValGluGluVallysSerProLysGluLysSerProGluLysAla   | 1730 |
| QY | 1242 | GAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAATAAACCTTCAAGGTTCCAAACCAAGA | 1301 |
| Db | 1731 | GluGluLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaAlaGlu  | 1750 |

|    |      |  |      |
|----|------|--|------|
| QY | 1302 | GAAGCTGACACTGAGTTGAGGAAG--CGGTCTCTCTGCTAAGGGAGGTGAGATTTCGATT   | 1358 |
| DB | 1751 | GluValLysSerProThrLysLysGluLysSerProGluLysSerAlaGluLysPro      | 1770 |
| QY | 1359 | CATTTCCAATTGAAGGAGGAGAGAGTCGAGCTGGAATGAATGATGTGAATGCCAAACGA    | 1418 |
| DB | 1771 | LysSerProThrLysLysGluSerSerProValLysMetAlaAspAspGluValLysSer   | 1790 |
| QY | 1419 | CCT-----GGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAATCTTAAGCAA         | 1466 |
| DB | 1791 | ProThrLysLysGluLysSerProGluLysValGluLysProAlaSerProThrLys      | 1810 |
| QY | 1467 | CATGGGAGAAAGGATTCTAAAATCACAGATCATTTTCATGAGAGTGCCTC-----        | 1514 |
| DB | 1811 | LysGluLysThrProGluLysSerAlaAlaGluLysSerProThrLysLysGlu         | 1830 |
| QY | 1515 | -----AAACGAGAGGACAAAAGAAAAGAAACAATGT--GAA                      | 1547 |
| DB | 1831 | LysSerProSerSerProThrLysLysThrGlyAspGluSerLysGluLysSerProGlu   | 1850 |
| QY | 1548 | ATGAAACATCAAGAACACAGAAAGGAGATCCCTAAATACATTCACCTCACCTTTCTCCA    | 1607 |
| DB | 1851 | LysProGluGluLysProLysSerProThrProLysLysSerProProGly--SerPro    | 1869 |
| QY | 1608 | GATAAGAAATGGCTTGGAACTCCTATTGAGGAGATGAGGAGAAATGCCAAGGTGTGGGATC  | 1667 |
| DB | 1870 | LysLysLysLysSerLysSerPro-----GluAlaGluLysProPro-----           | 1883 |
| QY | 1668 | CGCTGCCTCCCTTGAGACCATCTGCCAATCACACAGTGACTATTTCGGGTAGATCTTTTG   | 1727 |
| DB | 1884 | ---AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAla   | 1902 |
| QY | 1728 | CGAATAGGAGAAGTTCTTAACCTTTCCCAACACATTTTAAAGATTTTGTGGACAAACAAG   | 1787 |
| DB | 1903 | HisPhe---GluValValGluHisAlaThrGluCysLysTrpPheLeuAspGlyLys      | 1921 |
| QY | 1788 | CATGTTAAGATGCCTGTTCAGAACAAACTGTACCTCTGTGGAAGATGAGAATGGTGAG     | 1847 |
| DB | 1922 | GluIle-----ThrThrAlaGlnGlyValThrValSerLysAspAspGlnPheGlu       | 1938 |
| QY | 1847 | -----  | 1847 |
| DB | 1939 | PheArgCysSerIleAspThrThrMetPheGlySerGlyThrValSerValAlaSer      | 1958 |
| QY | 1848 | CGAGCTGCAGGCAGCCGGTGGAACTCATTCAGACTGCACCT---CTCAACAGGCTCACT    | 1904 |
| DB | 1959 | AsnAlaAlaGlySer-----ValGluThrLysThrGluLeuLysValLeuGlu          | 1974 |
| QY | 1905 | CGCCCCCAGAACCTGAAGATGCTATTCTGAAGTACAATGTGGCATATTCTTAAGAAATGG   | 1964 |
| DB | 1975 | ThrProLysGlu-----  | 1982 |
| QY | 1965 | GACTTTACA--GCTTTGATTGATTTC-----TGGGATAAGGTACTAGAAGAA           | 2009 |
| DB | 1983 | GluPheThrAspLysLeuArgAspMetGluValThrLysGlyAspThrValGlnMetAsp   | 2002 |
| QY | 2010 | GCAGAAGCTCAACAC-----TTGTATCAG-----                             | 2033 |
| DB | 2003 | ValIleAlaLeuHisSerProLeuTyrLysTrpTyrGlnAsnGlyAsnLeuLeuGluAsp   | 2022 |
| QY | 2034 | -----TCCATCTTGCCTGATATGGTGAAAATTGCACCTCTGTCTGCCAAAT            | 2078 |
| DB | 2023 | GlyLysAsnGlyValThrIleLysAsnGluGluAsnLysSerSerLeuIleIleProAsn   | 2042 |
| QY | 2079 | ATTTGTACCCAGCCAATACCACCTCTCTGAAAACAGAGATGAATCATTCATCACAAATGTCA | 2138 |
| DB | 2043 | -----AlaGlnAspSerGlyLysIleThrValGluAlaSerAsnGluValGly--Ser     | 2059 |
| QY | 2139 | CAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTCTGCACG-----              | 2183 |
| DB | 2060 | SerGluSerSerAlaGlnLeuThrValAsnProProSerThrThrProIleValValAsp   | 2079 |
| QY | 2184 | TTTCCACGA-----CGCAATGCCAAGATGAAATCAGAGTATTCC                   | 2222 |

Db 2080 GlyProLysSerValThrIleLysGluThrGluThrAlaGluPheLysAlaThrIleSer 2099  
|||::: |||::: |||::: |||::: |||:::  
QY 2223 AGTTATCCA-----GATATTAACTTCAATCGTTGTTTGAAGGACGT 2264  
::: ||| |||::: |||::: |||::: |||:::  
Db 2100 GlyPheProAlaProThrValLysTrpThrIleAsn---GluLysIleValGlu----- 2116  
|||::: |||::: |||::: |||::: |||:::  
QY 2265 TCATCAAGGAACACAGAGCTTAAACGCTCTTCTGCTACTTTAGAAGATCACAGAG 2324  
|||::: |||::: |||::: |||::: |||:::  
Db 2117 GluSerArgThrIleThrThrIleLysThrGluAspValThrLeuLysIleSerAsn 2136  
|||::: |||::: |||::: |||::: |||:::  
QY 2325 AAAAA-----CCCACTGGTTGGTGACATTCAAGACAGAGTCTT-----GAAGAT 2372  
|||::: |||::: |||::: |||::: |||:::  
Db 2137 AlaLysIleGluGlnThrGlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAsp 2156  
|||::: |||::: |||::: |||::: |||:::  
QY 2373 TTTCCAGAGTGGAA-----AGATGTGAAAAAATCCTGACTCGA 2411  
::: ||| |||::: |||::: |||::: |||:::  
Db 2157 SerLysGlnAlaAspLeuLysValGluProAsnValLysAlaProLysPheLysSerGln 2176  
|||::: |||::: |||::: |||::: |||:::  
QY 2412 CTG-----CATGTCACTTACGAAGGT-----ACCATAGAAGGA 2444  
|||::: |||::: |||::: |||::: |||:::  
Db 2177 LeuThrAspLysValAlaAspGluGlyGluProLeuArgTrpAsnLeuGluLeuAspGly 2196  
|||::: |||::: |||::: |||::: |||:::  
QY 2445 AACGGCCAGGGATGCTACAGGTGGATTTCGAAACCGTTTCGTTGGAGGTGGTGAACC 2504  
|||::: |||::: |||::: |||::: |||:::  
Db 2197 ProSerProGly----- 2200  
|||::: |||::: |||::: |||::: |||:::  
QY 2505 AGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTAAATCAACCTGAGTTGATTGTTTCA 2564  
|||::: |||::: |||::: |||::: |||:::  
Db 2201 -----ThrGluValSerTrpLeuLeuAsnGlyGlnProLeuThrLys 2214  
|||::: |||::: |||::: |||::: |||:::  
QY 2565 CGGCTCTTCACTGAGGTGCTGGATCACAATGAA---TGCTTTATCATCACAGGTACTGAG 2621  
|||::: |||::: |||::: |||::: |||:::  
Db 2215 SerAspThrValGlnValAspHisGlyAspGlyThrTyrHisValThrIleAlaGlu 2234  
|||::: |||::: |||::: |||::: |||:::  
QY 2622 CAGTACAGTGAATACACAGGC-----TATGCCGAACATACCGCTGGCGCGGAGCCAT 2675  
|||::: |||::: |||::: |||::: |||:::  
Db 2235 AlaLysProGluMetSerGlyThrLeuThrAlaLysAlaLysAsnAlaAla-GlyGluCy 2254  
|||::: |||::: |||::: |||::: |||:::  
QY 2676 GAAGACAGGAGCGAAAGGACGACTGGCAGAGGCGCACGACTGAGATCGTCCCATCGAC 2735  
::: ||| |||::: |||::: |||::: |||:::  
Db 2254 sGluThrSerAlaLysValThrValAsnGlyGlyAsnLysLysProGluPheValGlnAl 2274  
|||::: |||::: |||::: |||::: |||:::  
QY 2736 GCCCTCCACTTCAGACGCTACC 2757  
|||::: |||::: |||::: |||::: |||:::  
Db 2274 aProGlnAsnHisGluThrThr 2281  
|||::: |||::: |||::: |||::: |||:::

RESULT 15  
E85343  
hypothetical protein AT4g29440 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: E85343  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488; PMID:10617198  
A;Accession: E85343  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1071 <STO>  
A;Cross-references: GB:NC\_001268; NID:g7269843; PIDN:CAB79702.1; GSPDB:GN00140  
C;Genetics:  
A;Gene: AT4g29440  
A;Map position: 4

Alignment Scores:  
Pred. No.: 0.00137 Length: 1071  
Score: 170.00 Matches: 138  
Percent Similarity: 35.10% Conservative: 87  
Best Local Similarity: 21.53% Mismatches: 250  
Query Match: 2.35% Indels: 166  
DB: 2 Gaps: 30

US-09-302-812-1 (1-4070) x E85343 (1-1071)  
QY 45 GCTGTTAGGTGCCGGTGGAAAGGAAAGCAGGCGTCTGTAGTAGGGCCTGGTTCGGGAG 104  
|||::: |||::: |||::: |||::: |||:::  
Db 282 AlaAlaArgAlaAlaAlaGluSerAlaGluArgAlaSer-----Phe 295  
|||::: |||::: |||::: |||::: |||:::  
QY 105 GCTGTCAGAGCAGAGCTGCAGAAGCAGTCAGCGGCGAGAGGGGGCATGGTCCGGGAGGC 164  
|||::: |||::: |||::: |||::: |||:::  
Db 296 AlaAlaArgAlaAlaGluLeuSerSerLysGluArgMetMetMetMetGlnAsnSer 315  
|||::: |||::: |||::: |||::: |||:::  
QY 165 ACCGAG-----GAGGGGGCGCAGTCCGTCCTCCAGGGTTAGT 203  
|||::: |||::: |||::: |||::: |||:::  
Db 316 ThrGluSerArgAsnSerSerSerTyrGluAsnLeuArgSerAsnProProHis----- 333  
|||::: |||::: |||::: |||::: |||:::  
QY 204 GAATGAGGCTCTACGCCCGGCTGGCCCGGAGACTCAGTGTCTGCGGGTCCCAGCATGAGT 263  
|||::: |||::: |||::: |||::: |||:::  
Db 334 -----SerArgThrSerSerSerAsn---MetGlnGly 343  
|||::: |||::: |||::: |||::: |||:::  
QY 264 GCGGGCCCGGCTGTGAGCCCTGCACCAAG-----CGACCCCGCTGGGACCGCGCTGCA 317  
|||::: |||::: |||::: |||::: |||:::  
Db 344 GlyGlyPheGlyLysGluGluLeuLeuLysSerAsnAsnArgGlnValAspGlnSerThr 363  
|||::: |||::: |||::: |||::: |||:::  
QY 318 ACTTCTCCGCCCGCCTCGGACGCGCGGAGCTTC-----CCCGGCGAGGCAG 365  
|||::: |||::: |||::: |||::: |||:::  
Db 364 ThrThrThrArgAlaGluSerSerLysLysThrValAspGluLeuSerGluAsnThrSer 383  
|||::: |||::: |||::: |||::: |||:::  
QY 366 AGGCGCGTCTCGATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCGCTCCTCGTCA 425  
|||::: |||::: |||::: |||::: |||:::  
Db 384 TrpArgArgGlyHisSerArgGluAsnSerLeuGluMetArg-----ProAsnAspSer 401  
|||::: |||::: |||::: |||::: |||:::  
QY 426 GGCTGCGCCCTGGCGCGGCGGACAGCACCGGAGGCGCGCCACCTCTCTGTTTTCAAA 485  
|||::: |||::: |||::: |||::: |||:::  
Db 402 PheAlaLysIleGlyArgGluLysGlnGlnProGly-----MetAsp 415  
|||::: |||::: |||::: |||::: |||:::  
QY 486 CAGAAGACTATAACAGTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGT 545  
|||::: |||::: |||::: |||::: |||:::  
Db 416 AspIleAsnLeuSerSerSerAlaAspValLeuAsnLysLysGlnSerSerArgAlaSer 435  
|||::: |||::: |||::: |||::: |||:::  
QY 546 TTGCATAGTAAAGAAAAACAATACAAAGAGAAGAAATCCATGATGAT----- 593  
|||::: |||::: |||::: |||::: |||:::  
Db 436 SerHisSerProSerSerAsnPheSerAspAspAspValThrAlaLeuAspHisIle 455  
|||::: |||::: |||::: |||::: |||:::  
QY 594 -----TCTGTACAAAAAGATAACTTTTATCAACATAACATGGAATAATAGAAAAAT 644  
|||::: |||::: |||::: |||::: |||:::  
Db 456 AspSerProSerIlePheGluGluAsnLysPheGlnSerThrValGlyAspArgGluSer 475  
|||::: |||::: |||::: |||::: |||:::  
QY 645 GTTCTCAGTAGTTTGTATAAGTCAACAGTT----- 677  
|||::: |||::: |||::: |||::: |||:::  
Db 476 -----TyrAsnAspSerProValValValAlaProAlaPheAspAsp 490  
|||::: |||::: |||::: |||::: |||:::  
QY 677 ----- 677  
|||::: |||::: |||::: |||::: |||:::  
Db 491 TyrSerSerPheAspLysProGlnPheAspThrGluAspAlaTyrHisAspGluPro 510  
|||::: |||::: |||::: |||::: |||:::  
QY 678 GAAAAAGGTACACAGTAT-----TTGAAGCAGCATCAGACT 713  
|||::: |||::: |||::: |||::: |||:::  
Db 511 GluGlnGlyLeuGlyPheSerLeuLeuGlySerSerSerLysThrSerAspHisMetPro 530  
|||::: |||::: |||::: |||::: |||:::  
QY 714 GCGGCTATGTGTAAGTGGCAGAAATGAAGG-----CCA 746  
|||::: |||::: |||::: |||::: |||:::  
Db 531 ThrGluIleSerSerTrpSerLeuGluGlyHisLysAspLeuGlyLysLeuSerSerAla 550  
|||::: |||::: |||::: |||::: |||:::  
QY 747 CACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCAGCAGTTC 806  
|||::: |||::: |||::: |||::: |||:::  
Db 551 SerThrSerGlnValLeuGluLysGluLysProSer-----SerProProThrPhe 567  
|||::: |||::: |||::: |||::: |||:::  
QY 807 ACTAAT-----GCTAATGTCGATCAGTCGTCCTCCCA-----AAGGAT 842  
|||::: |||::: |||::: |||::: |||:::  
Db 568 AspAspGlyProThrSerProAlaSerLeuHisGluProGluProSerAlaLysPhe 587  
|||::: |||::: |||::: |||::: |||:::  
QY 843 GATCAGAGTGACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTGTGACACATGTA 902  
|||::: |||::: |||::: |||::: |||:::

Db 588 AspAspTyrAspArgAspSerGluSerGluGluAspAsnLeuGlyArgLeuSer----- 605  
QY 903 AAGCTTGCGAATGCAAAGCAGACGATGGAAGATGAACAGGCAGAGAAGCCAGAGCCAC 962  
Db 606 GlyArgAlaGluGlyLysSerLysLeuThrAlaGlnLysSerHisMetSerGluGlyPro 625  
QY 963 CAGAAAGTGTGGCAAGGCTTGCCATCCTGCA-----GAAAGCCTGTGCAGGGTGT 1010  
Db 626 AspAspLeuGlyArgTyrPhePheProSerAspThrGluAspGlnGlyAspAspSerLys 645  
QY 1011 CAGCAGGAGGAGACAGACGCTGCTGCCGAGAGCCCTTGTGCGACACTGGCTCTGAG--- 1067  
Db 646 ThrGlnGluGluSerAsp-----AlaGluThrPro-----ThrGlyLeuLysPhe 660  
QY 1068 ---GATGTTGGTACTGGACTGAAAATGCCAACAGATTGAATAGACAAGAAAGTAGTCTA 1124  
Db 661 GlyProLeuAlaSerGlyLeuGluAsnGluThrThrLeu-----ProSerTyr 676  
QY 1125 GGAATTTCTCCTCCATTTCAGAAAGAAAGTGAACCTGAG-----TCACCAATG 1172  
Db 677 GlySerSerProProArgAspLysThrSerSerLysSerLysLeuGlyLeuProThr 696  
QY 1173 GATGTAGATAATTCAAAATAGTTGTCCAGGATTCAGAAAGCAGATGAAGAGACAAGTCCA 1232  
Db 697 GluValAspProSerArgSerSerSerLeuGlnThrAlaSerSerSerSerIleArgAsn 716  
QY 1233 GGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAACAGCAATAAACCTTCAAGGTTTC 1292  
Db 717 GluLeuTyrThrGlnLysAlaSerAsnSerAspLys-----ArgProSerSerIle 733  
QY 1293 CAACCA-----AGAGAAGCTGACACTGAGTTGAGGAAGCGGTCTCT 1334  
Db 734 ProProAspSerSerSerAspAspGluSerAspMetGluLeuProLysArgValSer 753  
QY 1335 GCTAAGGGAGGTGAGATTCGATTACATTTCCAAATTTGAAGGAGGAGAGTCGAGCTGGA 1394  
Db 754 PheArgTyrGlnGluLysArgThrGluSerArgThrArgProThrHisLeuHisSerGly 773  
QY 1395 ATGAAT-----GATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAG 1448  
Db 774 ValSerHisLysAspLeuGluGluGluProThrArgAlaSerThrArgSerGlnAsp 793  
QY 1449 TGCAGAAATTTAAGCAACATGGGAGAAAGGATTCTAAATTCACAGATCATTTTCATGAGA 1508  
Db 794 ArgArgThrHisLysThrThrProAlaSerAlaSerAlaSerTyrPheHisThrMetSer 813  
QY 1509 GTGCCCCAAGCAGAGGACAAAAGAAAAGAAATGTGAAATGAAACATCAAAGAACA--- 1565  
Db 814 ---SerAspAspGluAspGluLysGluValHisArgAspThrAlaHisIleGlnThrArg 832  
QY 1566 -----GAAAGG----- 1571  
Db 833 ProTyrIleSerIleSerArgArgThrLysGlyGlnGluArgArgProSerLeuValThr 852  
QY 1572 ---AAGATCCCTAAATAC-----ATTCCACCTCACCTTTCTCCAGATAAG 1613  
Db 853 AlaLysIleAspLysValSerPheAspGluGluSerProProLysLeuSerProGluAla 872  
QY 1614 AAA 1616  
Db 873 Lys 873

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:32:43 ; Search time 34.05 Seconds  
(without alignments)  
12447.891 Million cell updates/sec

Title: US-09-302-812-1  
Perfect score: 7242  
Sequence: 1 accggaagtgaacgaagcc.....aaattttcattacaacaaaaa 4070

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xl  
-Q=/cgn2\_1/USPTO\_spool\_p/6333148/runat\_26052004\_150053\_5664/app\_query.fasta\_1.12437  
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=6333148 @CGN 1\_1\_128 @runat\_26052004\_150053\_5664 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description        |
|------------|-------|-------------|--------|--------------|--------------------|
| 1          | 184   | 2.5         | 3664   | 1 MINT_HUMAN | Q96t58 homo sapien |
| 2          | 180.5 | 2.5         | 1447   | 1 BUD4_YEAST | P47136 saccharomyc |
| 3          | 174.5 | 2.4         | 1462   | 1 NKCR_HUMAN | P30414 homo sapien |
| 4          | 174   | 2.4         | 1359   | 1 ATRX_CAEEL | Q9u7e0 caenorhabdi |
| 5          | 173   | 2.4         | 1210   | 1 AF4_HUMAN  | P51825 homo sapien |
| 6          | 171.5 | 2.4         | 1453   | 1 NKCR_MOUSE | P30415 mus musculu |
| 7          | 171   | 2.4         | 5147   | 1 PCLO_HUMAN | Q9y6v0 homo sapien |
| 8          | 170.5 | 2.4         | 6632   | 1 UN89_CAEEL | O01761 caenorhabdi |
| 9          | 169.5 | 2.3         | 913    | 1 IF38_HUMAN | Q99613 homo sapien |
| 10         | 168   | 2.3         | 1073   | 1 PVDA_PLAKN | P22545 plasmodium  |
| 11         | 164.5 | 2.3         | 646    | 1 SGL_BOVIN  | P23389 bos taurus  |
| 12         | 163   | 2.3         | 1070   | 1 PVDG_PLAKN | P50494 plasmodium  |
| 13         | 163   | 2.3         | 1362   | 1 BRD4_HUMAN | O60885 homo sapien |
| 14         | 162   | 2.2         | 513    | 1 DMP1_HUMAN | Q13316 homo sapien |
| 15         | 162   | 2.2         | 2158   | 1 MY9B_HUMAN | Q13459 homo sapien |
| 16         | 161.5 | 2.2         | 943    | 1 YM11_YEAST | P39523 saccharomyc |
| 17         | 161.5 | 2.2         | 5560   | 1 SPEN_DROME | Q8sx83 drosophila  |
| 18         | 160   | 2.2         | 633    | 1 MLH_TETTH  | P40631 tetrahymena |

|    |       |     |      |              |                    |
|----|-------|-----|------|--------------|--------------------|
| 19 | 158.5 | 2.2 | 5596 | 1 MDN1_HUMAN | Q9nu22 homo sapien |
| 20 | 158   | 2.2 | 1424 | 1 NCO3_HUMAN | Q9y6q9 h nuclear r |
| 21 | 158   | 2.2 | 3644 | 1 MINT_MOUSE | Q62504 mus musculu |
| 22 | 157   | 2.2 | 1411 | 1 TCOF_HUMAN | Q13428 homo sapien |
| 23 | 157   | 2.2 | 3122 | 1 DPOZ_MOUSE | Q61493 mus musculu |
| 24 | 156.5 | 2.2 | 773  | 1 ELAI_MOUSE | Q8cb77 mus musculu |
| 25 | 156   | 2.2 | 1341 | 1 ACIN_HUMAN | Q9ukv3 homo sapien |
| 26 | 155   | 2.1 | 489  | 1 DMP1_RAT   | P98193 rattus norv |
| 27 | 154.5 | 2.1 | 1338 | 1 ACIN_MOUSE | Q9jlx8 mus musculu |
| 28 | 154   | 2.1 | 406  | 1 SR40_YEAST | P32583 saccharomyc |
| 29 | 153.5 | 2.1 | 5085 | 1 PCLO_RAT   | Q9jks6 rattus norv |
| 30 | 153   | 2.1 | 5120 | 1 PCLO_CHICK | Q9pu36 gallus gall |
| 31 | 152.5 | 2.1 | 472  | 1 YWIE_CAEEL | Q23525 caenorhabdi |
| 32 | 152.5 | 2.1 | 915  | 1 CE05_HUMAN | Q9nyf5 homo sapien |
| 33 | 152   | 2.1 | 914  | 1 ORC1_YEAST | P54784 saccharomyc |
| 34 | 152   | 2.1 | 2004 | 1 MYS3_HUMAN | Q92794 homo sapien |
| 35 | 151   | 2.1 | 759  | 1 EPLI_HUMAN | Q9uhb6 homo sapien |
| 36 | 151   | 2.1 | 1311 | 1 ATRX_DROME | Q9ggn5 drosophila  |
| 37 | 150.5 | 2.1 | 503  | 1 DMP1_MOUSE | O55188 mus musculu |
| 38 | 150.5 | 2.1 | 1615 | 1 RIM1_RAT   | Q9jlr4 rattus norv |
| 39 | 150.5 | 2.1 | 2224 | 1 FA5_HUMAN  | P12259 homo sapien |
| 40 | 150.5 | 2.1 | 5038 | 1 PCLO_MOUSE | Q9qyx7 mus musculu |
| 41 | 150   | 2.1 | 771  | 1 CALD_CHICK | P12957 gallus gall |
| 42 | 149.5 | 2.1 | 1654 | 1 PCFB_HUMAN | Q94913 homo sapien |
| 43 | 149.5 | 2.1 | 1692 | 1 RIM1_HUMAN | Q86ur5 homo sapien |
| 44 | 149.5 | 2.1 | 2453 | 1 NCR1_MOUSE | Q60974 mus musculu |
| 45 | 149   | 2.1 | 845  | 1 NFM_RAT    | P12839 rattus norv |

ALIGNMENTS

RESULT 1  
MINT\_HUMAN STANDARD; PRT; 3664 AA.  
ID Q96T58; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;  
AC 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE Msx2-interacting protein (SMART/HDAC1 associated repressor protein).  
GN MINT OR SHARP OR KIAA0929.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,  
RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;  
RP RAR AND MTALL1.  
RC TISSUE=Liver, and Pituitary;  
RX MEDLINE=21231190; PubMed=11331609;  
RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,  
RA Hon M., Evans R.M.;  
RT "Sharp, an inducible cofactor that integrates nuclear receptor  
RT repression and activation.";  
RL Genes Dev. 15:1140-1151(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Bird C.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBBJ databases.  
RN [3]  
RP SEQUENCE OF 294-3664 FROM N.A.  
RA Rhodes S., Huckle E.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBBJ databases.  
RN [4]  
RP SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.  
RC TISSUE=Embryo, and Teratocarcinoma;  
RA Itoigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBBJ databases.





QY 207 TGAGGCTCTACGCCCG-----GGTGGCCCGGAGACTCAGT 242  
Dbb GluAlaLysGluProAlaGluThrLeuLysProProGluGlyTip--ArgSerProArg 1985  
QY 243 GCTGGGGTCCAGCATGAGTGGGGCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGC 302  
Dbb SerGlnLysThrAlaAlaGlyGlyProGlnGlyLysLysGlyLysAsnGluProLys 2005  
QY 303 TGGGACGCCGCTCAACTTCTCCGCCGGCGCTCGACGCCCGGAGCTTCCCGGCAGG 362  
Dbb ValAsp-----AlaThrArgProGluAlaThrThrGluVal-----GlyPro 2019  
QY 363 CAGAGCGCGCTCCTCGATTCCAAGGACGCTCCGGTGCAGTTTCAGGGTCCCGCTCTCG 422  
Dbb GlnIleGlyValLysGluSerSerMetGluPro-----LysAlaAlaGluGlu 2036  
QY 423 TCAGGCTGGCCCTGGCCGGCGGACAGCACCGGAGGCGCCACCTCTCTTGTTC 482  
Dbb AlaGlySerGluGlnLysArgAspArgLysAspAlaGly----- 2049  
QY 483 AAACAGAACTATTAACAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAA 542  
Dbb -----ThrAspLysAsnProProGluThrAlaProValGluValGluLysLys 2066  
QY 543 AGTTTGCATAGTAAGAAACAAACAATACAGAGAAGAAATCCATGATGAGTTCTGTACAA 602  
Dbb ProAlaProGluLysAsnSerLysSerLysArgGlyArgSerArgAsnSerArgLeuAla 2086  
QY 603 AAAGATAACTTTTATCAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGTTT 662  
Dbb ValAsp-----LysSerAlaSerLeuLysAsnValAspAlaAlaValSer 2101  
QY 663 GATAAGTCACCAAGTTGAAAAGGTACACAGATTTTGAAGCAGCATCAGACTCGGCTATG 722  
Dbb ProArgGlyAlaAlaAlaGlnAlaGlyGlu-----ArgGluSerGlyValValAlaVal 2119  
QY 723 TGTAAGTGGCAGATGAAGGCCACACTCAGAA-----CGGCTTTTGGAAAGT 770  
Dbb SerProGluLysSerGluSerProGlnLysGluAspGlyLeuSerSerGlnLeuLysSer 2139  
QY 771 GAACCTCCAGCGGTAACTCTGTGTACAGAGCAGTTCAGTAATGCTAATGTCATCAGTCG 830  
Dbb AspProValAspProAspLysGluProGluLys-----GluAspValSerAlaSer 2156  
QY 831 TCCCAAAGGATGATCAGATGACACAAATAGTGAGGAGATGAGATAATCAGCAGTTT 890  
Dbb GlyProSerProGlu-----AlaThrGlnLeuAlaLysGlnMetGluLeuGlnAla 2174  
QY 891 TTGACACATGTA--AAGCTTGCAATGCAAGCAGCAGCATGGAAGATGAACAGGCGCAGA 947  
Dbb ValGluHisIleAlaLysLeuAlaGluAlaSer----- 2185  
QY 948 GAAGCCAGAAGCCACCAGAAAGTGTGGCAAGGCTTGCCATCTGCAGAAGCCTGTGAGGG 1007  
Dbb -----AlaSerAlaAlaTyLysAlaAspAlaProGluGly 2197  
QY 1008 TGTACAGGAGGAGACAGACGTGGTGTCCGAGAGCCCTTGTCCGACACTGGCTCTGAG 1067  
Dbb LeuAlaProGluAspArgAsp-----LysProAlaHisGlnAlaSerGluThr 2213  
QY 1068 GATGTTGGTACTGGACTGAAAAT---GCCAACAGATTGAATAGACAAGAAAGTAGTCTA 1124  
Dbb GluLeuAlaAlaAlaIleGlySerIleAsnAspIleSerGlyGluProGluAsnPhe 2233  
QY 1125 GGAATTCCTCCTCCATTGTAGAAAAGAAAGTGAA----- 1157  
Dbb ProAlaProProTyProGlyGluSerGlnThrAspLeuGlnProProAlaGlyAla 2253  
QY 1158 -----CCTGAGTCACCAATGATGATAGATAATTCAAAAATAGTTGTGAG 1202  
Dbb GlnAlaLeuGlnProSerGluGluGlyMetGluThrAspGluAlaValSerGlyIleLeu 2273

QY 1203 GATTCAGAAAGCAGATGAAGAGACAAAGTCCAGGTTTTTGATGAACAGGAAGATAGCAGTTCT 1262  
Dbb GluThrGluAlaAlaThrGluSerSerArgProProValAsnAlaProAspProSerAla 2293  
QY 1263 GCTCAAAACAGCAAATAAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGG 1322  
Dbb GlyProThrAspThrLysGluAlaArgGlyAsnSerSerGluThrSerHisSerValPro 2313  
QY 1323 AAGCGTCTCTGCTAAGGAGGAGTGCAGATTGATTACATTTCATTTTGAAGGA----- 1376  
Dbb GluAlaLysGlySerLysGluValGluValThrLeuValArgLysAspLysGlyArgGln 2333  
QY 1376 ----- 1376  
Dbb LysThrThrArgSerArgArgLysArgAsnThrAsnLysLysValValAlaProValGlu 2353  
QY 1377 -----GGAGAGAGTCGAGCTCGAATGAATGATGTG 1406  
Dbb SerHisValProGluSerAsnGlnAlaGlnGlyGluSerProAlaAlaAsnGluGlyThr 2373  
QY 1407 AATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTTCTAAGCAA 1466  
Dbb ThrValGlnHisProGluAlaProGlnGlu-----GluLysGlnSerGluLysPro 2390  
QY 1467 CATGGGAGAAAGGATTTCTAAATTCACAGATCATTTTCATGAGAGTCCCAAGCAGAGGAC 1526  
Dbb HisSerThrProProGlnSerCysThrSerAspLeuSerLysIleProSerThrGluAsn 2410  
QY 1527 AAAAGAAAAGAAACAATGTGAATGAACATCAAAAGAACAGAAAGGAGATCCCTAAATAC 1586  
Dbb SerSerGlnGluIleSerValGluGluArgThrProThrLysAlaSerValProAsp 2430  
QY 1587 ATTCCACCTCACCTTCTCCAGATAAGAAATGGCTTGAACTCCTATTGAGCAGATGAGG 1646  
Dbb LeuProProProGlnPro-----AlaProValAspGlu----- 2442  
QY 1647 AGAATGCCAAGGTGTGGATCCGGCTG-----CCT 1676  
Dbb ---GluProGlnAlaArgPheArgValHisSerIleIleGluSerAspProValThrPro 2461  
QY 1677 CCCTTGAGACCATCT 1691  
Dbb ProSerAspProSer 2466

RESULT 2

BUD4\_YEAST STANDARD; PRT; 1447 AA.  
ID BUD4\_YEAST  
AC P47136;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Bud site selection protein BUD4.  
GN BUD4 OR YJR092W OR J1905.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9629555; PubMed=8707826;  
RA Sanders S.L., Herskowitz I.;  
RT "The BUD4 protein of yeast, required for axial budding, is localized  
RT to the mother/BUD neck in a cell cycle-dependent manner.";  
RL J. Cell Biol. 134:413-427(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ramezani Rad M., Kirchraht L., Hollenberg C.P.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Co-assembles with BUD3 at bud sites. BUD4 and BUD3 may  
CC cooperate to recognize a spatial landmark (the neck filaments)  
CC during mitosis and they subsequently become a landmark for  
CC establishing the axial budding pattern in G1.  
CC -!- SIMILARITY: Contains 1 PH domain.

```

CC CC -!- CAUTION: Ref.2 sequence differs from that shown in the N-terminal
CC CC section due to frameshifts.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U41641; AAB17116.1; -.
CC CC EMBL; Z49592; CAA89620.1; ALT_FRAME.
CC CC EMBL; Z49591; CAA89619.1; -.
CC CC PIR; S57113; S57113.
CC CC GermOnline; 141924; -.
CC CC SGD; S0003852; BUD4.
CC CC InterPro; IPR001849; PH.
CC CC Pfam; PF00169; PH; 1.
CC CC SMART; SM00233; PH; 1.
CC CC PROSITE; PS50003; PH_DOMAIN; 1.
CC CC ATP-binding.
CC CC KW DOMAIN 181 189 ASP/GLU-RICH (ACIDIC).
CC CC FT NP_BIND 1175 1182 ATP (POTENTIAL).
CC CC FT DOMAIN 1302 1413 PH.
CC CC FT CONFLICT 291 328 PTASVEDLNISNLPQADSSQNNPVTTDADALIENDVV ->
CC CC FT ANSLCGGFKHFNESSRVLIPAKIIQSLLMRRLKTMLC
CC CC FT (IN REF. 2).
CC CC FT CONFLICT 340 340 D -> E (IN REF. 2).
CC CC FT SEQUENCE 1447 AA; 164485 MW; 1D056B9FF1B7067B CRC64;
CC CC -----
CC CC Alignment Scores:
CC CC Pred. No.: 0.00142 Length: 1447
CC CC Score: 180.50 Matches: 184
CC CC Percent Similarity: 31.11% Conservative: 156
CC CC Best Local Similarity: 16.83% Mismatches: 362
CC CC Query Match: 2.49% Indels: 391
CC CC DB: 1 Gaps: 45
CC CC -----
CC CC US-09-302-812-1 (1-4070) x BUD4_YEAST (1-1447)
CC CC QY 465 GCCACCTCTCTGTTTTTCAAACAGAGACTATACCAGTTGGATGGACACTAAA----- 518
CC CC Db 417 SerThrSerThrLysSerGluGluSerTyrIleAlaAspTyrLysValThrArgGlnGlu 436
CC CC QY 519 -----GGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAAGAAAC----- 563
CC CC Db 437 AspTrpAspThrLysLysLeuHisGlnGluSerGluHisAlaAsnGluGlnProAlaIle 456
CC CC QY 564 -----AACAAATCAAGAGAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTTAT 617
CC CC Db 457 IleProGlnLysAspSerSerGluGluThrPheThrGluLeuAsnAsnGluSerGluPhe 476
CC CC QY 618 CAACATAACATG----- 629
CC CC Db 477 GlnArgAsnPheLysAspGlyGluGluTyrArgIleValGlnHisGluGluSerLeuTyr 496
CC CC QY 630 -----GAAAAAATTAGAAAATGTTTCTCAGCTAGGTTTGTAT 665
CC CC Db 497 GlyGlnArgThrLysSerProGluGluAsnIleIleAsnGlySerGluIleGlyValAsp 516
CC CC QY 666 -----AAGTCACCAAGTTGAAAAGGTACACAGTATTTGAAGCAG 704
CC CC Db 517 HisGlyGluAlaAlaGluValAsnGluProLeuAlaLysThrSer-----AlaGluGlu 534
CC CC QY 705 CATCAGACTCGGCTATGTGTAAGTGGCAG-----AATGAAGSGGCCACACTCAGAACGG 758
CC CC Db 535 HisAspLeuSerSerSerCysGluAspGlnSerValSerGluAlaArgAsnLysAspArg 554
CC CC QY 759 CTTTGTGGA----- 767
CC CC Db 555 IleGluGluLysGluValGluThrLysAspGluAsnIleGluThrGluLysAspGluSer 574

```







CC -|- SIMILARITY: Contains 1 cyclophilin-like ppiase domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L04288; AAA35734.2; -.  
DR EMBL; AF184110; AAD56402.1; -.  
DR PIR; A47328; A47328.  
DR HSP; Q27450; I433.  
DR Genew; HGNC:7833; NKTR.  
DR MIM; 161565; -.  
DR GO; GO:0016018; F:cyclosporin A binding; TAS.  
DR InterPro; IPR002130; CSA\_PPIase.  
DR Pfam; PF00160; pro isomerase; 1.  
DR PRINTS; PR00153; CSAPPISMRASE.  
DR PROSITE; PS00170; CSA\_PPIASE\_1; 1.  
DR PROSITE; PS50072; CSA\_PPIASE\_2; 1.  
KW Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.  
FT DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.  
FT DOMAIN 219 240 ARG/LYS-RICH (BASIC).  
FT DOMAIN 421 457 ARG/LYS-RICH (BASIC).  
FT DOMAIN 970 1010 ARG/LYS-RICH (BASIC).  
FT DOMAIN 194 244 ARG/SER-RICH.  
FT DOMAIN 466 574 ARG/SER-RICH.  
FT DOMAIN 664 814 ARG/SER-RICH.  
FT DOMAIN 1311 1348 ARG-SER TANDEM REPEAT-RICH.  
SQ SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;

Alignment Scores:  
Pred. No.: 0.00337 Length: 1462  
Score: 174.50 Matches: 158  
Percent Similarity: 33.84% Conservative: 110  
Best Local Similarity: 19.95% Mismatches: 298  
Query Match: 2.41% Indels: 226  
DB: 1 Gaps: 32

US-09-302-812-1 (1-4070) x NKCR\_HUMAN (1-1462)

QY 113 AGCAGAGTGCAGAAGCAGTCAAGCGGAGAGGGGCATGTGTCGGGAGGCACCGAGGA 172  
Db ThrHisSerSerArgAspSerTyrArgSerLysSerHisSerGlnSerTyrSerArgGly 531  
QY 173 GGGGGCGCAGTCCCTCCCGGTTAGTGAATGAGGCTTACGCCCGGCTGCCCG 232  
Db -SerSerArgSer-----ArgThrAlaSerLysSerSerHisSerArgSerAr 548  
QY 233 GAGACTCAGTGTGCGGGTCCCGATGATGAGTGGCGGCGCGGCTGTGAGCCCTGCACCAA 292  
Db gSerLysSerArg-----SerSerLysSerGlyHisArgLysArgAlaSerLy 565  
QY 293 GCGACCCCGTGGGACGCGCT-----GCAACTTCTCCGCC 328  
Db sSerProArgLysThrAlaSerGlnLeuSerGluAsnLysProValLysThrGluProLe 585  
QY 329 GCGCGCTCGGACGCCCGGAGC-----TCCCCCGGCGAGGAG 367  
Db uArgAlaThrMetAlaGlnAsnGluAsnValValGlnProValAlaGluAsnIl 605  
QY 368 GCGGTCCTCGATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCGCGCTCCTCGTCAGG 427  
Db eProValileProLeuSerAspSerProProSerArgTyrLysProGlyGlnLysPr 625  
QY 428 CTGC-----GCCCTGGCGCGGGGACAGCACCAGGCGCGCACCTCTCTGTGTTT 481  
Db oTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrHisLeuLeuPr 645  
QY 482 CAAACAGAGACTATAACCAAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGA 541

Db 645 oIleGlnSerThrTyr--SerLeuAlaAsnIleLys-----GluThrGlySerSerSe 662  
QY 542 AAGTTTGTCATAGTAAAGAAAAACAATACAGAGAAAGAAATCCATGATGAGTTCTGTATA-- 599  
Db rSerTyrHisLysArgGluLysAsnSerGluSerAspGlnSerThrTyrSerLysTyrSe 682  
QY 600 -----CAAAAAGATAAAGTTTA 616  
Db rAspArgSerSerGluSerSerProArgSerArgSerArgSerSerArgSerArgSerTy 702  
QY 617 TCAACATACATGGAAAAATTAGAAAATGTT---TCTCAGCTAGGTTTGTGATAAGTCACC 673  
Db rSerArgSerTyrThrArgSerArgSerLeuAlaSerSerHisSerArgSerArgSerPr 722  
QY 674 AGTTGAAAAAGGT-----ACACAGTATTGAGCAGCATCAGACTGCGGCTATGTG 724  
Db oSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArgSerSe 742  
QY 725 TAAAGTGG-----CAGAATGAAGGGCCACACTCAGAAACGGCTTTTGGAAAGTGA 772  
Db rSerTyrThrSerIleSerSerAspAspGlyArgArgAlaLysArgArgLeuArgSerSe 762  
QY 773 ACCTCCAGCGGTAACCTCGGTACCAGAGCAGTTCAGTAATGCTAATGTC----- 821  
Db rGlyLysLysAsnSerValSerHisLysLysHisSerSerSerSerSerGluLysThrLeuHi 782  
QY 822 -----GATCAGTCGTCCCCAAAGGATGATCAGTGCACAGTGCACACAAA 859  
Db sSerLysTyrValLysGlyArgAspArgSerSerCysValArgLysTyrSerGluSerAr 802  
QY 860 TAGT-----GAGGAGGTAGAGATAATCAGCAGTTCACACATGTAAAGCTTGCGAA 913  
Db gSerSerLeuAspTyrSerSerAspSerGluGln-----SerSerValGlnAlaThrGl 820  
QY 914 TGCAAAGCAGACGATGGAAAGATGAACAGGGCAGAGAGCCAGAACCCACCAGAGTGIG 973  
Db nSerAlaGlnGluLysGlu---LysGlnGlyGlnMetGluArgThrHisAsnLysGlnGl 839  
QY 974 CAAGGCTTGCCATCCTGCAGAAAGCCTGTGCAGGGTGTGCAGGAGGAGACAGACGTGGT 1033  
Db uLys-----AsnArgGlyGluGluLysSerLysSerGl 850  
QY 1034 GTCGAGAGCCCTTGTGCGAC-----ACTGGCTCTGAGGATGTGGTACTGGACTGAA 1087  
Db uArgGluCysProHisSerLysLysArgThrLeuLysGluAsnLeuSerAspHisLeuAr 870  
QY 1088 AAATGCCAACAGATTGAATAGACAAAGTAGTCTAGGAAATTCCTCCATTTGAGAA 1147  
Db gAsnGlySerLysProLysArgLys-----AsnTyrAlaGlySerLysTr 885  
QY 1148 AGAAAGTGAACCTGAGTCAACCAATGGATGTA--GATAATCCAAAAATAGTTGTCAGGA 1204  
Db pAspSerGluSerAsnSerGluArgAspValThrLysAsnSerLysAsnAspSerHisPr 905  
QY 1205 TTCAGAGCAGATGAAGAGACAAAGTCCAGGTTTGTGTAACAGGAAGATAGCAGTTCTGC 1264  
Db oSerSerAspLysGluGluGlyGluAlaThrSerAspSerGluSerGluValSerGluIl 925  
QY 1265 TCAAAACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAAGCTGACACTGAGTTGAGAA 1324  
Db eHisIleLysValLysProThr-----Th 933  
QY 1325 GCGGTCTCTGCTAAGGGAGGTGAGATTTCGATTACATTTCCAATTTGAAGGAGGAGAGAG 1384  
Db rLysSerSer-----Th 937  
QY 1385 TCGAGCTGGAATGAATGATGTGAAT-----GCCAAACGACCTGGAAGTAC 1429  
Db rAsnThrSerLeuProAspAspAsnGlyAlaTrpLysSerSerLysGlnArgThrSerTh 957  
QY 1430 TTCTAGCCTGAATGTAGTGCAGAAATTTCTAAG----- 1463  
Db rSerAspSerGluGlySerCysSerAsnSerGluAsnAsnArgGlyLysProGlnLysHi 977

|    |      |   |      |
|----|------|---|------|
| QY | 1464 | -CAACATGGGAGAAAGGATTCTTAAAAATCACAGATCATTTTCATGAGAGTGCCCAAAGCAGA | 1522 |
| Db | 977  | sLysHisGlySerLysGluAsnLeuLysArgGluHis-----ThrLysLysVally        | 994  |
| QY | 1523 | GGACAAAAGAAAACAACAAATGTGAAATGAAACATCAAAAGAACAGAAAGGAGATCCCTAA   | 1582 |
| Db | 994  | sGluLysLeuLysGlyLysLysAspLysLysHisLysAlaProLysArgLysGlnAlaPh    | 1014 |
| QY | 1583 | ATACATTCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACCTCCTATTGAGGAGAT    | 1642 |
| Db | 1014 | eHisTrpGlnPro-----  | 1018 |
| QY | 1643 | GAGGAGAATGCCAGGTGTGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAATCACAC      | 1702 |
| Db | 1018 | -----   | 1018 |
| QY | 1703 | AGTGACTATTTCGGGTAGATCTTTTGGGAATAGGAGAAGTTCTTAAACCTTTCCCAACACA   | 1762 |
| Db | 1019 | -----ProLeuGluPheGlyGluG1                                       | 1025 |
| QY | 1763 | TTTTAAAGATTGTGGGACAAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAACCTTGTA   | 1822 |
| Db | 1025 | uGluGluGluGluIleAspAspLysGlnValThrGlnGluSerLysGluLysLysVal--    | 1044 |
| QY | 1823 | CCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACCTCATTCAGAC   | 1882 |
| Db | 1045 | -----SerGluAsnAsnGlu-----                                       | 1049 |
| QY | 1883 | TGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAGTACAA   | 1942 |
| Db | 1050 | -----ThrIleLysAspAsnIleLeuLysThrG1                              | 1059 |
| QY | 1943 | TGTGGCA-----TATTCTAAGAAATGGGACTTTACAGCTTTTGATTGATTCTGTG         | 1990 |
| Db | 1059 | uLysSerSerGluGluAspLeuSerGlyLysHisAspThrValThrValSerSerAspLe    | 1079 |
| QY | 1991 | GGATAAGGTACTAGAGAAGCAGAAGCTCAACACTTGATCAGTCCATCTTGCCTGATAT      | 2050 |
| Db | 1079 | uAspGlnPheThrLysAspAspSerLysLeuSerIleSerProThrAlaLeuAsnThrG1    | 1099 |
| QY | 2051 | GGTGAAAATTTGCACTCTGTCTGCCAAAATATTGTATCCAGCCCAATACCACCTCCTGAAACA | 2110 |
| Db | 1099 | uGluAsnValAla--CysLeuGlnAsnIle-----G1                           | 1109 |
| QY | 2111 | GAAGATGAATCATTCATCACAATGTCAACAGGAACAGATTGCCAGTCTTTTAGCTAATGC    | 2170 |
| Db | 1109 | nHisValGluGluSerValProAsnGlyValGluAspValLeuGlnThrAspAspAsnMe    | 1129 |
| QY | 2171 | TTTCTTCTGCAGTTTCCAGCAGCAATGCCAAGATGAATCAGAGTATTCAGT-----        | 2225 |
| Db | 1129 | tGluIleCysThr---ProAspArgSerSerProAlaLysValGluThrSerProLe       | 1148 |
| QY | 2226 | -----TATCCAGATATTAAC  | 2240 |
| Db | 1148 | uGlyAsnAlaArqLeuAspThrProAspIleAsn                              | 1159 |

Villard L., Fontes M., Ewbank J.J.;

"Characterization of xnp-1, a Caenorhabditis elegans gene similar to the human XNP/ATR-X gene.";

Gene 236:13-19(1999).

[2]

SEQUENCE FROM N.A.

STRAIN=Bristol N2;

Fulton R., Wohldmann P.;

Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

-!- FUNCTION: Could be a global transcriptional regulator. Modifies gene expression by affecting chromatin (Potential).

-!- SUBCELLULAR LOCATION: Nuclear.

-!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

US-09-302-812-1 (1-4070) x ATRX_CAEEL (1-1359)

QY 516 AAAGGAATCAAGCACGTTGAATCAGAAAGTTTGCATAGTAAAGAAAACAACAATACAAGA 575
    |||| |||| :::||||| ||| |||:::
Db 81 LysSerArgLysArgAlaLysSerGluSer---GluSerAspGluSerAspGluGlu 99
    |||| |||| :::||||| ||| |||:::
QY 576 GAAGAAATCCATGATGAGTTCGTACAAAAAGATAACTTTTATCAACATAACATGGA AAAA 635
    ::||| ||| :::||| ::| ::|
Db 100 AspArgLysLysSerLysSerLysLysValAspGlnLysLysGluLysSerLys 119
    ::||| ||| :::||| ::| ::|
QY 636 TTAGAAAATGTTTCTCAGCTAGGTTTGTATAAGTCACCAAGTTGAAAAAGGTACACAGTAT 695
    ::||| ::| |||::| |||::|
Db 120 LysLysArgThrThrSerSerSerGluAspGluAspSerAspGluGluArgGluGln--- 138
    ::||| ::| |||::| |||::|
QY 696 TTGAAGCAGCATCAGACTCCGGCTATGTGTAAAGTGGCAGAATGAAGGCCACACTCAGAA 755
    |||| ::| ::| ||| |||
Db 139 ---LysSerLysLysSerLysLysThrLysLysGlnThrSerSerGluSerSerGlu 157
    |||| ::| ::| ||| |||

```

|  |                         |
|--|-------------------------|
| RESULT 4   |                         |
| ATTRX_CAEEEL   |                         |
| ID ATTRX_CAEEEL  | STANDARD; PRT; 1359 AA. |
| AC Q9U7E0; O02061;   |                         |
| DT 28-FEB-2003 (Rel. 41, Created)  |                         |
| DT 28-FEB-2003 (Rel. 41, Last sequence update)                           |                         |
| DT 15-MAR-2004 (Rel. 43, Last annotation update)                         |                         |
| DE Transcriptional regulator AttrX homolog (X-linked nuclear protein-1). |                         |
| GN XNP-1 OR B0041.7.   |                         |
| OS Caenorhabditis elegans.   |                         |
| OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  |                         |
| OC Rhabditidae; Peloderinae; Caenorhabditis.                             |                         |
| OX NCBI_TaxID=6239;  |                         |
| RN [1]   |                         |
| RP SEQUENCE FROM N.A., AND CHARACTERIZATION.                             |                         |
| RX MEDLINE=93365296; PubMed=10433961;                                    |                         |

```
QY 756 CGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCAGAGCAGCTTCAGTAATGCT 815
Db 158 -----GluSerGluGluGluArgLysValLysLysSerLysLysAsnLysGluLys 174
QY 816 AATGTCGATCAGTCTCCCAAGGATGATCAGCTGACACAAATAGTAGGAGAGTAGA 875
Db 175 SerValLysLysArgAlaGluThrSerGluGluSerAspGluAspGluLysProSerLys 194
QY 876 GATAATCAGCAGTTTGTGACACATGTAAGCTTGCGAATGCAAGCAGCAGATGGAAGAT 935
Db 195 LysSerLysLysGlyLeuLys-----LysLysAlaLysSerGluSerGluSer 212
QY 936 GAACAGGGCAGAGAGCCAGAGCCACCAGAGTGTGGCAAGGCTTGCCATCCTGCAGAA 995
Db 213 GluAspGluLysGluValLysLysSerLysLysLysSerLysValValLysLysGlu 232
QY 996 GCCTGTGCAGGGTGTACAGCAGGAGGAGACAGACGCTGGTGTCCGAGAGCCCTTGTCCGAC 1055
Db 233 SerGluSerGluAspGluAlaProGluLysLysLysThrGluLysArgLysArgSerLys 252
QY 1056 ACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAACAGATTGAATAGACAAGAA 1115
Db 253 ThrSerSerGluGlu-----SerSerGlu 260
QY 1116 AGTAGTCTAGGAATCTCCTCCATTTTGAGAAAGAGTGAACCT-----GAG 1163
Db 261 SerGluLysSerAspGluGluGluGluGluLysGluSerProLysProLysLysLys 280
QY 1164 TCACCAATGGATGATGATAAATTCACAAAATAGTTGTCAGGATTGAGAGCAGATGAAGAG 1223
Db 281 LysProLeuAlaValLysLysLeuSerSerAspGluGluSerGluGluSerValGlu 300
QY 1224 ACAAGTCCAGGT-----TTTGATGAACAGCAGAGATAGC 1256
Db 301 ValLeuProGlnLysLysLysArgGlyAlaValThrLeuIleSerAspSerGluAspGlu 320
QY 1257 AGTTCTGCTCAACACAGCAATAAACCTTCAAGGTTCCAAACAGAGAGCTGACACTGAG 1316
Db 321 LysAspGlnLysSerGluSerGluAlaSerAspValGluGluLysValSerLysLysLys 340
QY 1317 TTGAGGAAGCGTCTGCTAAGGGAGGTGAGATTCCATTACATTTCCTCAATTGAAGGA 1376
Db 341 AlaLysLysGlnGluSerSerGlu-----SerGly 350
QY 1377 GGAGAGAGTCGAGCTGGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGC 1436
Db 351 SerAspSerSerGluGlySerIleThrValAsnArgLys-----Ser 364
QY 1437 CTGAATGTAGAGTGCAGAAATTTCTAAGCAACATGGG-----AGAAAGGATTCTAAATC 1490
Db 365 LysLysLysGluLysProGluLysLysLysLysGlyIleIleMetAspSerSerLysLeu 384
QY 1491 ACAGATCATTTTCATGAGAGTGTCCCAAGCAGAGGACAAAAGAAAAGAACAAATGTGAAATG 1550
Db 385 GlnLysGluThrIleAspAlaGluArgAlaGluLysGluArgArgLysArgLeuGluLys 404
QY 1551 AAACATCAAGA-----1562
Db 405 LysGlnLysGluPheAsnGlyIleValLeuGluGluGluAspLeuThrGluMetLeu 424
QY 1563 -----ACAGAAAGGAGATCCCTAAATACATTCACCTCACCTTTCTCCAGAT 1610
Db 425 ThrGlyThrSerSerGlnArgLysLeuLysSerValVal-----LeuAspProAsp 441
QY 1611 AAGAAATGGCTTGGAACTCCTATTGAGGAGATGAGGAGATGCCAAGGTGTGGGATCCGG 1670
Db 442 -----SerSerThrValAspGluGluSerLysLysPro-----452
QY 1671 CTGCCTCCCTTGAGACCATCTGCCAATCACACAGTACATTCGGGTAGATCTTTTGGGA 1730
Db 453 -----ValGluValHisAsnSerLeuValArg 461
QY 1731 ATAGGAGAAGTTCCTAAACCTTTTCCCAACACAT---TTTAAAGATTGTGGGACAACAAG 1787
```

```
Db 462 Ile-----LeuLysProHisGlnAlaHisGlyIleGlnPheMetTyrAsp-----476
QY 1788 CATGTTAAGATGCCTTGTTCAGAACAAAACTTGTACCTCTGTGGAAGATGAGAATGGTGAG 1847
Db 477 -----CysAlaCysGluSerLeuAspArgLeuAspThrGluGly-----489
QY 1848 CGAGCTGCAGCAGCCGGTGGAACTCATTCAGACTGCACCTTCTCAACAGGCTCAGCTCGG 1907
Db 489 -----489
QY 1908 CCCCAGAACCTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCTAAGAAAATGGGAC 1967
Db 490 -----SerGlyGlyIleLeuAlaHisCysMetGlyLeuGlyLysThr-----503
QY 1968 TTTACAGCTTTGATTGATTCTCTGGGATAAGGTACTAGAAAGACAGAAAGCTCAACACTTG 2027
Db 504 ---LeuGlnValIleThrPheLeuHisThrValLeu-----Met 515
QY 2028 TATCAGTCCATCTTGCCTGATATGGTGAAAATTCACACTCTGTCTGCCA---AATATTGT 2084
Db 516 HisGluLysIleGlyGluLysCysLysArgValLeuValValProLysAsnValIle 535
QY 2085 ACCAGCCAATACCACCTCCTGAAAACAGAAAGATGAATCATTCATCACAATGTCACAGGAA 2144
Db 536 IleAsnTrpPhe-----LysGluPheGlnLysTrpLeuValAspAsnAspGlu 551
QY 2145 CAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGACGCTTCCACGACGCAATGCCAAG 2204
Db 552 GluLeuAspThrIleAspValAsn-----559
QY 2205 ATGAAATCAGAGTATTCAGTTATCCAGATATTAACATTCATCGG-----TTGTTTGA 2258
Db 560 -----GluLeuAspSerTyrLysThrIleGluAspArgArgAlaLeuLysAla 576
QY 2259 GGACGTTTCATCAAGGAAACCAGAGAGCTTAAACGCTCTTCTGTACTTTAGAAGAGTC 2318
Db 577 TrpHisSerSerLysThrProSerValMetIleIleGlyTyrAspLeuPheArgIleLeu 596
QY 2319 ACA-----GAGAAAAACCCACTGGTGGTGACATTCAACAAGACAG 2360
Db 597 ThrValGluAspAspProLysLysLysLysProLysAsnArgAsnArgLeuGluLys 616
QY 2361 AGCTTTGAAGATTTT-----CCAGAGTGGGAAAGATGTGA 2396
Db 617 AlaLysGluAspPheArgLysTyrLeuGlnAsnProGlyProAspMetValValCysAsp 636
QY 2397 -----AAACTCCTG 2405
Db 637 GluAlaHisLysLeuLysAsnAspAspSerAlaLeuSerLysCysMetValLysIleLeu 656
QY 2406 ACTCGACTGCATGTCACCTTACGAAGGTACC---ATAGAAGGAAAC-----2447
Db 657 ThrLysArgArgIleCysLeuThrGlyThrProLeuGlnAsnAsnLeuMetGluTyrHis 676
QY 2448 -----GGCCAGGGCATGCTA-----CAGGTGGATTTTGCAAAACCGT 2483
Db 677 CysMetValAsnPheValLysProGlyLeuLeuGlyThrLysThrGluPheAlaAsnArg 696
QY 2484 TTCGTTGGAGGTGTAAACCAGTGCAGGACTTGTGCAAGAA-----GAAATC 2531
Db 697 PheVal-----AsnIleIleAsnArgGlyArgThrLysAspAlaSerProLeuGluVal 714
QY 2532 CGCTTTTGA 2540
Db 715 SerPheMet 717
```

```
RESULT 5
AF4 HUMAN
ID AF4 HUMAN STANDARD; PRT; 1210 AA.
AC P51825;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
```







QY 931 AAGATGAACAGGCGAGAGAGCCAGAGAGCCAGAGAGTGTGGCAAGGCTTGCCATCCTG 990  
Db 506 luProProThrThrAsnLysTrpGlnLeuAspAsnTrpLeuThrLysValSerGlnProA 526  
QY 991 CAGAAGCCTGTGCGGGTGTGAGCAGGAGGAGAGACAGCTGGTGTCCGAGAGCCCTTGT 1050  
Db 526 laAlaProProGluGly-----ProArgSerThrGluProProArgArgHisProGluS 544  
QY 1051 CGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAACAGATTGAATAGAC 1110  
Db 544 er---LysGlySerSerAspSerAlaThrSer-----G 554  
QY 1111 AAGAAAGTAGTCTAGGAAATCTCTCTCCATTGTGAGAAAAGAAAGTGAA---CCTGAGTCAC 1167  
Db 554 lngluHisSerGluSerLysAspProProProLysSerSerSerLysAlaProArgAlap 574  
QY 1168 CAATGGATGTAGATAAT--TCCAAAAATAGTTGTTCAGGATTTCAGAGCAGATGAAGAGA 1224  
Db 574 roProGluAlaProHisProGlyLysArgSerCysGlnLysSerProAla----- 590  
QY 1225 CAAGTCCAGGTTTGTATGAACAGGAAGATAGCAGTTCTGTCTCAACAGCAATAAACCTT 1284  
Db 591 -----GlnGlnGluProProGlnArgGlnThrValGlyThrLysG 604  
QY 1285 CAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCTAAGGGAG 1344  
Db 604 lnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSerLeuGlnGly- 623  
QY 1345 GTGAGATTTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATGAATGATG 1404  
Db 624 -----GluArgGluProGlyLeuLeuProTyrGlySerArgAspG 637  
QY 1405 TGAATGCCAAA---CGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAATTCTA 1461  
Db 637 lnThrSerLysAspLysPro-----LysValL 646  
QY 1462 AGCAACATGGGAGA--AAGATTCTAAATACAGATCATTTTCATGAGAGTGCCCAAAG 1518  
Db 646 ysThrLysGlyArgProArgAlaAlaSerAsnGluProLysProAlaValProProS 666  
QY 1519 CAGAGGACAAAAGAAAAGAACAAATGT----- 1544  
Db 666 erSerGluLysLysLysHisLysSerSerLeuProAlaProSerLysAlaLeuSerGlyP 686  
QY 1545 --GAAATGAACATCAAAGAACAGAAAGGAGATCCCTAAATACATT----- 1589  
Db 686 roGluProAlaLysAspAsnValGluAspArgThrProGluHisPheAlaLeuValProL 706  
QY 1590 -----CCACCTCACCTTTCTCCAGATAAAGAAATGGCTTGA----- 1625  
Db 706 euThrGluSerGlnGlyProProHisSerGlySerGlySerArgThrSerGlyCysArgG 726  
QY 1626 -----ACTCCTATTGAGGAGATGAGGAGAAATGCCAAGGTGTGGGATC----- 1667  
Db 726 lnAlaValValGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspT 746  
QY 1668 -----CGGCTGCCCTTGAGA-----CCATCTGCCAATCACACAGTGACTATTTC 1713  
Db 746 hrLysLeuLeuSerProLeuArgAspThrProProGlnSerLeuMetValLysIleT 766  
QY 1714 GGGTAGATCTTTTTCGGAATAGGAGAAGTTCTTAAACCTTTCCCAACACATTTTAAAGATT 1773  
Db 766 hrLeuAspLeuLeu-----SerArgIleProGlnPro----- 776  
QY 1774 TGTGGGACAAACAGCATGTTAAGATGCCTTGTTCAGAAACAAAACCTTGTTACCTGTG---G 1830  
Db 777 -----ProGlyLysGlySerArgGlnArgLysAlaGluAspLysGlnProProAlaGlyL 795  
QY 1831 AAGATGAGAATGGTGAGCGAGCTGCAGGAGCCCGGTGGGAACCTCATTCAGACTGCACCTTC 1890  
Db 795 ysLysHisSerSerGluLysArgSerSerAspSerSerLysLeuAlaLysLysArgL 815

QY 1891 TCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAGTACAATGTGGCAT 1950  
Db 815 ysGlyGluAlaGluArgAspCysAspAsnLysLysIleArgLeuGluLysGluIleLysS 835  
QY 1951 ATTCTAAGAAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAGGTACTAGAAGAAG 2010  
Db 835 erGlnSerSerSerSerSerHisLysGluSerSerLysThrLysProSerArgp 855  
QY 2011 CAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGTTGGTGAATAATGCACCTCTGC 2070  
Db 855 roSerSerGlnSerSerLysLysGluMetLeuProPro----- 867  
QY 2071 TGCCAAATATTGTACC-----CAGCCAATA---CCACTCTGAAACAGAAATGA 2118  
Db 868 --ProProValSerSerSerSerGlnLysProAlaLysProAlaLeuLysArgSerArgA 887  
QY 2119 ATCATTCATCAACAATGTACAGGAA-----CAGATTGCCAGTCTTTTAGCTAATG 2169  
Db 887 rgGluAlaAspThrCysGlyGlnAspProProLysSerAlaSerSerThrLysSerAsnH 907  
QY 2170 CTTTCTCTGCACGTTTCCACGA-----CGCAATGCCAAGATGAAAT 2211  
Db 907 isLysAspSerSerIleProLysGlnArgValGluGlyLysGlySerArgSerS 927  
QY 2212 CAGAGTATTCC----- 2222  
Db 927 erGluHisLysGlySerSerGlyAspThrAlaAsnPropheProValProSerLeuProA 947  
QY 2223 -----AGTTATCCAGATATTAACTTCAATCGGTTTGAAGGACGTT 2265  
Db 947 snGlyAsnSerLysProGlyLysProGlnValLysPheAspLysGln-GlnAlaAspLeu 966  
QY 2266 CATCAAGGAAACCAGAGAAGCTTAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGA 2325  
Db 967 HisMetArgGluAlaLysLysMetLysGlnLysAlaGluLeuMet----- 981  
QY 2326 AAAAAACCCTGGTGGTGACATTCAACAAGACAGAGTCTTGAAGATTTTCCAGAGTGGG 2385  
Db 982 -----ThrAspArgValGlyLysAlaPheLysTyrLeu 992  
QY 2386 AAAGATGTGAAAAACTCCTGACTCGACTGCATGTCACCTTACGAAGGTACCATAGAAGAA 2445  
Db 993 GluAlaValLeu---SerPheIleGluCysGlyIleAlaThrGluSerGlu---SerGln 1010  
QY 2446 ACGGCCAGGCATGCTACAGGTGGATTGTGCAACCGTTTCGTTG----- 2490  
Db 1011 SerSerLysSerAlaTyrSerValTyrSerGluThrValAspLeuIleLysPheIleMet 1030  
QY 2491 -----GAGGTGGTGTAAACCAAGTCAGGAC----- 2514  
Db 1031 SerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlaValLeu 1050  
QY 2515 -----TTGTGCAAGAAGAAAATCCGCT 2535  
Db 1051 CysMetArgCysGlnSerIleLeuAsnMetAlaMetPheArgCysLysLysAspIleAla 1070  
QY 2536 -----TTTTAATCAACCTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATC 2589  
Db 1071 IleLysTyrSerArgThrLeuAsnLysHisPheGluSerSerSerLysValAlaGlnAla 1090  
QY 2590 ACAATGAATGCTTATCATCACAGGTACTGACAGTACAGTGAATACACAGGCTATGCCG 2649  
Db 1091 ProSerProCysIleAlaSerThrGlyThrProSerProLeuSerProMetProSerPro 1110  
QY 2650 AAACATACCGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGACGACTGGCAGAGGC 2709  
Db 1111 AlaSerSerValGly---SerGlnSerSerAlaGlySerValGlySerSerGlyValAla 1129  
QY 2710 GCACGACTGAGATCGTCGCCATCGACGCCCTCCACT-----TCAGAC 2751  
Db 1130 AlaThrIle-----SerThrProValThrIleGlnAsnMetThrSerSerTyr 1145  
QY 2752 GCTACCTCGACCAGTTTGTGCCCGAGAAAGATCAGAC-----GGGAGCTTA 2796

Db 1146 ValThrIleThrSerHisValLeuThrAlaPheAspLeuTrpGluGlnAlaGluAlaLeu 1165  
QY 2797 ACAAGGCTTACTGTGGATTCTTC 2820  
Db 1166 ThrArgLysAsnLysGluPhePhe 1173

RESULT 6  
NKCR\_MOUSE STANDARD; PRT; 1453 AA.  
AC P30415;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-related protein) (NK-TR protein).  
GN NKTR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93133824; PubMed=8421688;  
RA Anderson S.K.; Gallinger S., Roder J., Frey J., Young H.A., Ortaldo J.R.;  
RT "A cyclophilin-related protein involved in the function of natural killer cells";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).  
RN [2]  
RP REVISIONS TO C-TERMINUS.  
RC STRAIN=BALB/c; TISSUE=Blood;  
RA Anderson S.K.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Component of a putative tumor-recognition complex.  
CC Involved in the function of NK cells.  
CC -!- SIMILARITY: Contains 1 cyclophilin-like PPIase domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; L04289; AAA37500.2; ALT\_INIT.  
DR HSSP; Q27450; 1A33.  
DR MGD; MGI:97346; Nktr.  
DR InterPro; IPR002130; CSA\_PPIase.  
DR Pfam; PF00160; pro\_isomerase; 1.  
DR PRINTS; PR00153; CSAPPISMRASE.  
DR PROSITE; PS00170; CSA\_PPIASE\_1; 1.  
DR PROSITE; PS50072; CSA\_PPIASE\_2; 1.  
KW Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.  
FT DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.  
FT DOMAIN 222 241 ARG/LYS-RICH (BASIC).  
FT DOMAIN 422 459 ARG/LYS-RICH (BASIC).  
FT DOMAIN 964 1003 ARG/LYS-RICH (BASIC).  
FT DOMAIN 198 273 ARG/SER-RICH.  
FT DOMAIN 468 565 ARG/SER-RICH.  
FT DOMAIN 658 812 ARG/SER-RICH.  
FT DOMAIN 1303 1453 ARG-SER TANDEM REPEAT-RICH.  
SQ SEQUENCE 1453 AA; 163439 MW; DF1173FF814B283E CRC64;

Alignment Scores:  
Pred. No.: 0.00517 Length: 1453  
Score: 171.50 Matches: 193  
Percent Similarity: 30.91% Conservative: 117  
Best Local Similarity: 19.24% Mismatches: 384  
Query Match: 2.37% Indels: 309  
DB: 1 Gaps: 40

US-09-302-812-1 (1-4070) x NKCR\_MOUSE (1-1453)  
QY 101 GGAGGCTGTCAAGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGGCATGGTCCCGG 160  
Db 508 GlySerAlaSerThrHisSerSerArgAspSerTyrArgSerLysSerHisSerArgSer 527  
QY 161 AGGCACCGAGGAGGGGC-----GCAGTCCGTCCTCCCTCCAGGGTTAGTGAATG--- 208  
Db 528 AspSerArgGlySerSerArgSerArgAlaValSerLysSerSerSerArgSerLeuAsn 547  
QY 209 -----AGGCTCTACGCCCGGGCTGGCCCGGAGACTCAGTGTGCGGGTCCAG 256  
Db 548 ArgSerLysSerArgSerSerArgSerGlyPro-ArgArgThrSerIleSerProLy 567  
QY 257 CATGAGTGCGGGCCCGGC-----TGTGAGCCCTGCACCAAGCGACC 298  
Db 567 sLysProAlaGlnLeuSerGluAsnLysProValLysThrGluProLeuArgProSerVa 587  
QY 299 CCGCTGGGACGCCGCTGCAACTTCTCCGCCGCCGCTCGGACGCCCGGAGCTTCCCGCG 358  
Db 587 lProGlnAsnGlyAsnValLeuValGlnProValAla-----AlaGluAsnIlePro-- 604  
QY 359 CAGGCAGAGCGCGCTCCTCGATTCCAAGGACGCTCCGGTGCAGTTCCAGGGTCCCGCGTC 418  
Db 605 -----ValleProLeuSerAspSerProProSerArgTrpLysProGl 620  
QY 419 CTCGTCAAGCTGC-----GCCCTGGCGCGGGGACAGCACCGAGCGCGCCACCTC 472  
Db 620 yGlnLysProTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrThrHi 640  
QY 473 TCTTGTTTTCAAAACAGAGACTATAACAGTTGGATGGACACTAAAGGAATCAAG---AC 529  
Db 640 sLeuLeuProValGlnSerThrTyrSer-----LeuThrAsnIleLysAlaTh 656  
QY 530 AGTTGAATCAGAAAGTTTGCATAGTAAAGAAACAAACAATACAAGAGAAGATCCATGAT 589  
Db 656 rValSerSerSerTyrHisLysArgGluLysProSerGluSerAspGlySerAlaTy 676  
QY 590 GAGTTCTGTACAAAAGATAACTTTTATCAACATAACATGGAATAATTAGAAAATGTTC 649  
Db 676 rSerLysTyrSer----- 680  
QY 650 TCAGCTAGGTTTGTGATAAGTCACCAGTTGAAAAGGTACACAGTATTTTGAAG----- 701  
Db 681 -----AspArgSerSerGlySerSerGlyArgSerGlySerLysSerSerAr 696  
QY 702 ----CAGCATCAGACTCGCGCTATGTGTAAGTGCAGAAATGAAGGGCCACACTCAGAACG 757  
Db 696 gSerArgSerSerArgSerTyrThrArgSerArgSerArgSerLeuProThrSerAr 716  
QY 758 GCTTTTGGAAAGTGAACCTCCAGCGGTAACCTCTGTGTACCAGAGCAGTTTCAGTAATGCTAA 817  
Db 716 gSerLeuSerArgSerProSerSerArgSerHisSerProAsnLysTyrSerAspGlySe 736  
QY 818 TGTCGATCAGTCGTCCCAAGGATGATCACAGTCAGACACAAATAGTGAGGAGAGTAGA-- 875  
Db 736 r-----GlnHisSerArgSerSerTyrThrSerValSerSerAspGlyArgAr 754  
QY 876 -----GATAATCAGCAGTTTTTGACACATGTAAAG-----CTTCCGAA 913  
Db 754 gAlaMetPheArgSerAsnArgLysLysSerValThrSerHisLysArgHisArgSerAs 774  
QY 914 TGCAAAGCAGACGATGGAAGATGAA-----CAGGCAGAGAGCCAGAACG---CACCA 964  
Db 774 nSerGluLysThrLeuHisSerLysTyrValArgGlyArgGluLysSerSerArgHisAr 794  
QY 965 GAAGTGTGGCAAGGCTTGCCTATCTGCAGAAAGCCCTGTGCAGGGTGTTCAGCAGGAGGAGAC 1024  
Db 794 gLysTyrSerGluSerArgSerSerLeuAspTyrThrSerAspSerAspGlnSerHisVa 814  
QY 1025 AGACGTGGTGTCCGAGAGGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACT 1084  
Db 814 lGlnValTyrSerAlaProGluLysGluLysGlnGlyLysValGluAlaLeuAsnAspLy 834









Db 1169 LysSerGlnValGlnIle---AlaGluGluLysLeuGluGlyArg-----ValAlaPro 1185  
QY 159 GGAGGCACCGAGGAGGGCGCAGTCCGTCCTCCAGGTTAGTGAATGAGGCTCTACG 218  
Db 1186 LysThrValGlnGluGlyLysGln-----ProGlnThrLysMetGluGluLeuPro 1202  
QY 219 CCCGGCTGCCCGGAGACTC-----AGTGTCTCGGGTCCCAGCATGAGTGC 266  
Db 1203 SerGlyThrProGlnSerLeuProLysGluAspAspLysThrThrLysThrIleLysGlu 1222  
QY 267 GGCCCGGCTGTGAGCCCTGCACCAAGCGACCC-----CGCTGG 305  
Db 1223 GlnProGln---ProProCysThrAlaLysProAspGlnGluLysGluAspAspLysSer 1241  
QY 306 GACGCCGCTGCAACTTCTCCCGCGCGCCCTCGGACCGCCCGGAGCTTCCCGCGCAGGCAG 365  
Db 1242 AspThrSerSerSerGlnGlnProLysSerProGlnGlyLeuSerAspThrGlyTyr--- 1260  
QY 366 AGGCGCGTCTCGATTCCAAGGACGCTCCGGTGCAAGTTCAGGGTCCCGCGCTCCTCGTCA 425  
Db 1261 -----SerSerAsp 1263  
QY 426 GGCTGCGCCCTGGCGCGGGCGGAGCAGCACCGAGCGCCACCTCTCTGTGTTTCAAA 485  
Db 1264 GlyIleSer-----SerSerLeuGlyGluIleProSerLeuIle----- 1276  
QY 486 CAGAAGACTATAACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGT 545  
Db 1276 ----- 1276  
QY 546 TTGCATAGTAAAGAAACAAACAATACAAGAGAAGAAATCCATGATGAGTTCTGTACAAAAA 605  
Db 1277 -----ProThrAspGluLysAspIleLeuLysGlyLeuLysLys 1289  
QY 606 GATAACTTTTATCAACAT-----AACATCGAAAAAATTAGAAAT 644  
Db 1290 AspSerPheSerGlnGluSerSerProSerSerProSerAspLeuAlaLysLeuGluSer 1309  
QY 645 -----GTTTCTCAGCTAGGTTTGTGATAAGTACCAGTT 677  
Db 1310 ThrValLeuSerIleLeuGluAlaGlnAlaSerThrLeuAlaAspGluLysSer----- 1327  
QY 678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTCGGGTATGTGTAAGTGGCAGAAT 737  
Db 1328 GluLysLysThrGln-----ProHisGluValSerPro----- 1338  
QY 738 GAAGGCCACACTCAGAACCGCTTTTGGAA-----AGTGAACCTCCAGCGGTAACCTG 791  
Db 1339 GluGlnProLysAspGlnGluLysThrGlnSerLeuSerGluThrLeuGluIleThrIle 1358  
QY 792 GTACCAGAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTC-----CCAAAGGATGAT 845  
Db 1359 SerGluGluGluIleLysGluSerGlnGluGluArgLysAspThrPheLysLysAspSer 1378  
QY 846 CACAGTGAC---ACAAATAGTACGAGAGTAGAGATAATCAGCAGTTTGTGACACATGTA 902  
Db 1379 GlnGlnAspIleProSerSerLysAspHisLysGluLysSerGluPheValaspIle 1398  
QY 903 AAGCTTGGC-----AATGCAAAGCAGACGATGGAAGATGAACAGGGC--- 944  
Db 1399 ThrThrArgArgGluProTyrAspSerValGluGluSerSerGluSerGluAsnSerPro 1418  
QY 945 -----AGAGAAGCCAGAAAGCCACAGAAAGTGTGGCAAGCTTGCCATCCTGCGAGAA 995  
Db 1419 ValProGlnArgLysArgThrSerValGlySerSerSerSerAspGluTyrLysGln 1438  
QY 996 GCCTGTGCGGGTGTACAGCAGGAGGAGACAGACGTGGTGTCCGAGAGCCCTTGTCCGAC 1055  
Db 1439 GluAspSerGlnGlySerGlyGluGluGluAspPheIleArgLysGlnIleIleGluMet 1458  
QY 1056 ACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAACAGATTGAATAGACAAGAA 1115  
::: ||||| ::::| ::::| ::::|

Db 1459 SerAlaAspGluAsp---AlaSerGlySerGluAspAspGluPheIleArgAsnGlnLeu 1477  
QY 1116 AGTAGTCTAGAAATTTCTCTCCATTTGAGAAAAAGTGAACCTGAG----- 1163  
Db 1478 LysGluIleSerSerSerThrGluSerGlnLysLysGluGluThrLysGlyLysGlyLys 1497  
QY 1164 -----TCACCAATGGATGTAGAT 1181  
Db 1498 IleThrAlaGlyLysHisArgArgLeuThrArgLysSerSerThrSerIleAspGluAsp 1517  
QY 1182 AATTCCAAAAATAGTTGTTCAGGATTTCAGAACCATGATGAAGAGACAAGTCCAGGTTTGTAT 1241  
Db 1518 AlaGlyArgArgHisSerTrpHisAspGluAspAspGlu-----AlaPheAsp 1533  
QY 1242 GAACAGGAAGATAGCAGTTCTGTCTCAACACAGCAAAATAAACCTTCA----- 1286  
Db 1534 GluSerProGluLeuLysTyrArgGluThrLysSerGlnGluSerGlnGluLeuValVal 1553  
QY 1287 -----AGGTTCCAAACCAAGAGAAAGCTGACACTGAGTTGAGGAAG 1325  
Db 1554 ThrGlyGlyGlyLeuArgArgPheLysThrIleGluLeuAsnSerThrIleAlaAsp 1573  
QY 1326 CGGTCTCTCTAAGGAGGTGAGATTTCGATTACATTTCCAATTT----- 1370  
Db 1574 LysTyrSerAlaGluSerSerGlnLysLysThrSerLeuTyrPheAspGluGluProGlu 1593  
QY 1371 -----GAAGGAGGAGAG 1382  
Db 1594 LeuGluMetGluSerLeuThrAspSerProGluAspArgSerArgGlyGluGlySerSer 1613  
QY 1383 AGTCGAGCTGGAATGAATGATGTGAATGCCRAACGACCTGGAAGTACTTTCTAGCCTGAAT 1442  
Db 1614 SerLeuHisAlaSerSerPheThrProGlyThrSerProThrSerValSerSerLeuAsp 1633  
QY 1443 GTAGAGTGCAGAAATTTCTAAGCAACATGGGAGAAAGGATTTCTAAA----- 1487  
Db 1634 GluAspSerAspSerSerProSerHisLysLysGlyGluSerLysGlnArgLysAla 1653  
QY 1488 ---ATCACAGATCATTTTCATGAGAGTGCCTCAAGAGCAGAGACAAAA----- 1529  
Db 1654 ArgHisArgProHisGlyProLeuLeuProThrIleGluAspSerSerGluGluGlu 1673  
QY 1530 ---AGAAAAGAACAAATGTGAAATGAAACATCAAGAAACAGAAAGGAAGATC----- 1577  
Db 1674 LeuArgGluGluGluGluLeuLysGluGlnGluLysGlnArgGluIleGluGlnGln 1693  
QY 1577 ----- 1577  
Db 1694 GlnArgLysSerSerSerLysLysSerLysLysAspLysAspGluLeuArgAlaGlnArg 1713  
QY 1578 -----CCTAAATACATTTCCACCTCCACCTTTCTCCAGATAAGAAATGCGCTTGA 1625  
Db 1714 ArgArgGluArgProLysThrProProSerAsnLeuSerProIleGluAsp-----Ala 1731  
QY 1626 ACTCCTATTGAGGAGATGAGGAGAAATGCCAAGGTGTGGGATCCGGCTGCCCTTGTGAGA 1685  
Db 1732 SerProThrGluGluLeuArgGln-----AlaAlaGluMetGluGluLeuHis 1747  
QY 1686 CCATCT-----GCCAATCACACAGTACTATTCCGGGTAGATCTTTTGGGAATAGGAGAA 1739  
Db 1748 ArgSerSerCysSerGluTyrSerProSerIleGluSerAsp----- 1761  
QY 1740 GTTCTTAAACCTTTTCCCAACACATTTTAAAGATTTGTGGGAC---AACAAAGCATGTTAAG 1796  
Db 1762 ---ProGluGlyPheGluIleSerProGluLysIleIleGluValGlnLysValTyrLys 1780  
QY 1797 ATGCCTTGTTCAGAACAAAACCTTGTAACCTGTGGAAGATGAGAATGGTGAGCGAGCTGCA 1856  
Db 1781 LeuProThrAlaValSerLeuTyrSerProThrAspGluGlnSerIleMetGlnLysGlu 1800  
QY 1857 GGCAGCCGGTGGGAACCTCAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAAC 1916  
::: ||||| ::::| ::::|  
Db 1801 GlySerGln-----LysAla 1805



```
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGEY -> RRRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960E89 CRC64;
```

Alignment Scores:  
Pred. No.: 0.00931 Length: 6632  
Score: 170.50 Matches: 198  
Percent Similarity: 35.12% Conservative: 142  
Best Local Similarity: 20.45% Mismatches: 418  
Query Match: 2.35% Indels: 210  
DB: 1 Gaps: 47

US-09-302-812-1 (1-4070) x UN89\_CABEL (1-6632)

```
QY 246 GCGGTCCAGCATGAGTCGGGCCCCGGCTGTGAG---CCCTGCACCAAGCAGCCCGC 302
Db 1394 AlaSerProThrLysLysThrGlyGluGluValLysSerProLysGluLysSerProAla 1413
QY 303 TGGGACGCGCTGCAACTTCTCCGCGCGCGCTCGGACGCCGGAGCTTCCCGGCAGG 362
Db 1414 SerProThrLysLysGluLysSerProAlaAlaGluGluValLysSer---ProThrLys 1432
QY 363 CAGAGGCGGTCCTCGATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCGTCCTCG 422
Db 1433 LysGluLys-----SerProSerProThrLysLysGluLysSerProLysSer 1449
QY 423 TCAGGCTGCGCCCTGGCGCGGGGACAGCACCGCAGGCGCCACCTCTCTGTTTTC 482
Db 1450 ProThrLysLysThrGlyAspGluValLysGluLysSerProProLysSerProThrLys 1469
QY 483 AAACAGAAGACTATAACCACTTGGATGGACACTAA----- 518
Db 1470 LysGluLysSerProGluLysProGluAspValLysSerProValLysLysGluLysSer 1489
QY 519 -----GGAATCAAGACAGTTGAATCAGAA---AGTTGCATAGTAAGAAAC 563
Db 1490 ProAspAlaThrAsnIleValGluValSerSerGluThrThrIleGluLysThrGluThr 1509
QY 564 AACAAATACAAGAGAA-----GAATCCATGATGAGTTCTGTACAAAAAGAT 608
```

```
Db 1510 ThrMetThrThrGluMetThrHisGluSerGluGluSerArgThrSerValLysLysGlu 1529
QY 609 AACTTTTATCAACATAACATGGAATAATTAGAAAATGTTTCTCAGCTAGGTTTGATAAG 668
Db 1530 LysThrProGluLysValAspGluLysProLysSerProThrLysLys-----AspLys 1547
QY 669 TCACCAGTTGAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTCGCGCTATGTGTAAAG 728
Db 1548 SerProGluLysSerIleThrGluGluIleLysSerProValLys----- 1562
QY 729 TGGCAGATGAAGGGCCACACTCAGAACCGCTTTTGAAAGTGAACCTCCAGCGGTAAC 788
Db 1563 ---LysGluLysSerProGluLysValGluGluLysProAlaSerProThrLysLysGlu 1581
QY 789 CTGGTACCAGACAGCTTCAGTAATGCTAATGTGATCGATCAGTCGTCGCCAAAAGGATGATCAC 848
Db 1582 LysSerProGluLys-----ProAlaSerProThrLysLysSerGluAsnGluValLys 1599
QY 849 AGTGACACAAATAGTCAGGAGAGTAGAGATAATCAGCAGTGTTCGACACATGTAAAGCTT 908
Db 1600 SerProThrLysLysGluLysSerProGlu---LysSerValValGluGluLeuLysSer 1618
QY 909 GCGAATGCAAGCAGACGATGGAA---GATGAACAGGGCAGAGAGCCAGAACCCACCAG 965
Db 1619 ProLysGluLysSerProGluLysAlaAspAspLysProLysSerProThrLysLysGlu 1638
QY 966 AAGTGTGGCAAGGCTTGCCATCCTGCAGAACCCCTGTGCAGGCTGTACAGGAGGAGACA 1025
Db 1639 LysSerProGluLysSerAlaThrGluAspValLysSerProThrLysLysGluLysSer 1658
QY 1026 ---GACGTGGTGTCCGAGAGCCCTTGTGCGACACACTGGCTCTGAGGATGTTGGTACT--- 1079
Db 1659 ProGluLysValGluGluLysProThrSerProThrLysLysGluSerSerProThrLys 1678
QY 1080 -----GGACTGAAAAATGCCAACAGATTGAATAGACAGAAAGTAGTCTAGGA 1127
Db 1679 LysThrAspAspGluValLysSerProThrLysLysGluLysSerProGlnThrValGlu 1698
QY 1128 AATTCTCCT-----CCATTGTGAAAAAGTGAACCTGAGTCACCAATGATGATAGAT 1181
Db 1699 GluLysProAlaSerProThrLysLysGluLysSerProGlu----- 1712
QY 1182 AATTCGAAAAATAGTTGTCAGGATTCAGAAGCAGATGAAGAGACAGTCCAGGTTTGTAT 1241
Db 1713 -----LysSerValValGluGluValLysSerProLysGluLysSerProGluLysAla 1730
QY 1242 GAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAATAAACCTTCAAGGTCCAAACCAAGA 1301
Db 1731 GluGluLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaAlaGlu 1750
QY 1302 GAAGCTGACACTGAGTTGAGGAAG---CGGTCTCTGCTAAGGAGGAGTGAGATTCGATTA 1358
Db 1751 GluValLysSerProThrLysLysGluLysSerProGluLysSerAlaGluGluLysPro 1770
QY 1359 CATTTCCAAATTTGAAGGAGGAGAGAGTCGAGCTGGAATGAATGATGTAATGCCAAACGA 1418
Db 1771 LysSerProThrLysLysGluLysSerProValLysMetAlaAspAspGluValLysSer 1790
QY 1419 CCT-----GGAAGTACTTCTAGCTGAATGTAGAGTCAGAGTGCAGAAATCTAAGCAA 1466
Db 1791 ProThrLysLysGluLysSerProGluLysValGluGluLysProAlaSerProThrLys 1810
QY 1467 CATGGGAGAAAGGATTCTAAATCACAGATCATTTTCATGAGAGTGCCC----- 1514
Db 1811 LysGluLysThrProGluLysSerAlaAlaGluGluLeuLysSerProThrLysLysGlu 1830
QY 1515 -----AAAGCAGAGGACAAAAGAAAAGAACAAATGT---GAA 1547
Db 1831 LysSerProSerSerProThrLysLysThrGlyAspGluSerLysGluLysSerProGlu 1850
QY 1548 ATGAAACATCAAAAGAACAGAAAGAGATCCCTAAATACATTCACCTCACCTTCTCCA 1607
```



Db 1851 LysProGluGluLysProLysSerProThrProLysLysSerProProGly---SerPro 1869  
QY 1608 GATAAGAAATGGCTTGGAACTCCTATTGAGGAGATGAGGAGAATGCCAAGGTGGGATC 1667  
Db 1870 LysLysLysLysSerLysSerPro-----GluAlaGluLysProPro----- 1883  
QY 1668 CGGCTGCCTCCCTTGAGACCATTCTGCCAATCACACAGTGACTATTCCGGGTAGATCTTTTG 1727  
Db 1884 ---AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAla 1902  
QY 1728 CGAATAGGAGAAGTTCCTAAACCTTTCCCAACACATTTTAAAGATTTGTGGGACAACAAG 1787  
Db 1903 HisPhe---GluValValGluHisAlaThrGluCysLysTrpPheLeuAspGlyLys 1921  
QY 1788 CATGTTAAGATGCCTTGTTCAGAACAAACTTGTACCTGTGTGGAAGATGAGAATGGTGAG 1847  
Db 1922 GluIle-----ThrThrAlaGlnGlyValThrValSerLysAspAspGlnPheGlu 1938  
QY 1847 ----- 1847  
Db 1939 PheArgCysSerIleAspThrThrMetPheGlySerGlyThrValSerValValAlaSer 1958  
QY 1848 CGAGCTGCAGGCAGCGGTGGAACTCATTCAGACTGCACTT---CTCAACAGGCTCACT 1904  
Db 1959 AsnAlaAlaGlySer-----ValGluThrLysThrGluLeuLysValLeuGlu 1974  
QY 1905 CGGCCCCAGAACCTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCTAAGAAATGG 1964  
Db 1975 ThrProLysGlu-----ThrLysLysPro 1982  
QY 1965 GACTTTACA---GCTTTGATTGATTC-----TGGGATAAGGTACTAGAAGAA 2009  
Db 1983 GluPheThrAspLysLeuArgAspMetGluValThrLysGlyAspThrValGlnMetAsp 2002  
QY 2010 GCAGAAGCTCAACAC-----TTGTATCAG----- 2033  
Db 2003 ValIleAlaLeuHisSerProLeuTyrLysTrpTyrGlnAsnGlyAsnLeuLeuGluAsp 2022  
QY 2034 -----TCCATCTTGCCTGATATGGTGAATAATGCACCTCTGTCTGCGCAAAT 2078  
Db 2023 GlyLysAsnGlyValThrIleLysAsnGluGluAsnLysSerSerLeuIleIleProAsn 2042  
QY 2079 ATTTGTACCCAGCCAATACCACTCCTCGAAACAGAGATGAATCATTCATCACAAATGTCA 2138  
Db 2043 -----AlaGlnAspSerGlyLysIleThrValGluAlaSerAsnGluValGly---Ser 2059  
QY 2139 CAGGAACAGATTGCCAGTCTTTAGCTAATGCTTTCTTCGCACG----- 2183  
Db 2060 SerGluSerSerAlaGlnLeuThrValAsnProProSerThrThrProIleValValAsp 2079  
QY 2184 TTTCCACGA-----CGCAATGCCAAGATGAATCAGAGTATTCC 2222  
Db 2080 GlyProLysSerValThrIleLysGluThrGluThrAlaGluPheLysAlaThrIleSer 2099  
QY 2223 AGTTATCCA-----GATATTACTCAATCGGTTGTTTGAAGGACGT 2264  
Db 2100 GlyPheProAlaProThrValLysTrpThrIleAsn---GluLysIleValGlu----- 2116  
QY 2265 TCATCAAGGAAACCAGAGAAGCTTAAACGCTCTTTCTGCTACTTTTAGAAGAGTCACAGAG 2324  
Db 2117 GluSerArgThrIleThrIleLysThrGluAspValTyrThrLeuLysIleSerAsn 2136  
QY 2325 AAAAAA-----CCCACTGGTGGTGACATTCACAAGACAGAGTCTT-----GAAGAT 2372  
Db 2137 AlaLysIleGluGlnThrGlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAsp 2156  
QY 2373 TTTCCAGAGTGGAA-----AGATGTGAAAAA---CTCTGACTCGA 2411  
Db 2157 SerLysGlnAlaAspLeuLysValGluProAsnValLysAlaProLysPheLysSerGln 2176  
QY 2412 CTG-----CATGTCACTTACGAAGGT-----ACCATAGAAGGA 2444  
Db 2177 leuThrAspLysValAlaAspGluGlyGluProLeuArgTrpAsnLeuGluLeuAspGly 2196

QY 2445 AACGGCCAGGCGATGCTACAGGTGGATTTTGCAAAACCGTTTCGTTGGAGGTGGTGTAAAC 2504  
Db 2197 ProSerProGly----- 2200  
QY 2505 AGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTAAATCAACCCCTGAGTTGATTGTTTCA 2564  
Db 2201 -----ThrGluValSerTrpLeuLeuAsnGlyGlnProLeuThrLys 2214  
QY 2565 CGGCTCTTCACTGAGGTGCTGGATCACAAATGAA---TGTCTTATCATCACAGTACTGAG 2621  
Db 2215 SerAspThrValGlnValValAspHisGlyAspGlyThrTyrHisValThrIleAlaGlu 2234  
QY 2622 CAGTACAGTGAATACACAGGC-----TATGCCGAAACATACCGCTGGGCCCGAGCCAT 2675  
Db 2235 AlaLysProGluMetSerGlyThrLeuThrAlaLysAlaLysAsnAlaAla-GlyGluCy 2254  
QY 2676 GAAGACAGGAGCGAAGGACGACTGGCAGAGGCGCACGACTGAGATCGTCGCCATCGAC 2735  
Db 2254 sGluThrSerAlaLysValThrValAsnGlyGlyAsnLysLysProGluPheValGlnAl 2274  
QY 2736 GCCCTCCACTTCAGACGCTACC 2757  
Db 2274 aProGlnAsnHisGluThrThr 2281

RESULT 9

IF38\_HUMAN  
ID IF38 HUMAN STANDARD; PRT; 913 AA.  
AC Q99613; O00215;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Eukaryotic translation initiation factor 3 subunit 8 (eIF3 p110)  
DE (eIF3C).  
GN EIF3S8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97150873; PubMed=8995409;  
RA Asano K., Kinzy T.G., Merrick W.C., Hershey J.W.B.;  
RT "Conservation and diversity of eukaryotic translation initiation  
RT factor eIF3.";  
RL J. Biol. Chem. 272:1101-1109(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99425270; PubMed=10493829;  
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,  
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,  
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,  
RA Eichler E.B., Harris P.C., Venter J.C., Adams M.D.;  
RT "Genome duplications and other features in 12 Mb of DNA sequence from  
RT human chromosome 16p and 16q.";  
RL Genomics 60:295-308(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,  
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,





```
Db      362 AlaGluAsnLeuGlyGluGlyValIleValIleLysPheAsnIleAlaSer 381
QY      1956 AAGAAATGGGACTTTACAGCTTTGATT-----GATTTCTGGGATAAG 1997
Db      382 ---LeuTyrAspTyrAsnProAsnLeuAlaThrTyrMetLysProGluMetTyrGlyLys 400
QY      1998 GTACTAGAAAGACAGAAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTAAA 2057
Db      401 CysLeuAsp-----CysIleAsnGluLeuMetAsp 410
QY      2058 ATTGCACCTCTCTGTGCCAAATATT----- 2081
Db      411 IleLeuPheAlaAsnProAsnIlePheValGlyGluAsnIleLeuGluGluSerGluAsn 430
QY      2082 -----TGTACCCAGCCCAATACCCTC-----CTGAACACAGAAG 2114
Db      431 LeuHisAsnAlaAspGlnProLeuArgValArgGlyCysIleLeuThrIleuValGluArg 450
QY      2115 ATGAATCATCTCCATCACAATG-----TCACAGGAACAGATT 2150
Db      451 MetAspGluGluPheThrLysIleMetGlnAsnThrAspProHisSerGlnGluTyrVal 470
QY      2151 GCCAGTCTTTAGCTAATGCTTTCTTCTGACAGCTTCCACGACGCAATGCCAAGATGAAA 2210
Db      471 GluHisLeuLysAspGluAlaGlnValCysAlaIleIleGluArgValGlnArgTyrLeu 490
QY      2211 TCAGAGTATTCAGATTATCCAGATATTAATCAATCGGTTGTTTGAAGACGTTTCAT-- 2268
Db      491 GluGluLysGlyThrThrGluGluVal-CysArgIleTyrLeuLeuArgIleLeuHisTh 510
QY      2269 -----CAAGGAAACCAGAGAAGCTTA-----AAACGCTCTTCTGTC 2303
Db      510 rTyrTyrLysPheAspTyrLysAlaHisGlnArgGlnLeuThrProProGluGlySerSe 530
QY      2304 TACTTTAGAAGAGTCACAGAGAAAAAACCCTGCTGGTGGTGACATTCACAAGACAGAGT 2363
Db      530 rLysSerGluGlnAspGlnAlaGluAsn-----GluGlyGluAspSerAl 545
QY      2364 CTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACTCCTGACTCGACT----- 2413
Db      545 aValLeuMet-GluArgLeuCysLysTyrIleTyrAlaLysAspArgThrAspArgIleA 565
QY      2414 -----GCATGTCACTTACGAAGGTACCATAGAAAGAAACGCCAGGGCATG 2459
Db      565 rgThrCysAlaIleLeuCysHisIle-----TyrHis-----HisA 577
QY      2460 CTACAGGTGGATTTTGCAAAACCGTTTCGTTGGAGGTGGTGTAAACCAGTGCAGGACTT--- 2516
Db      577 la-----LeuHisSerArgTrp-TyrGlnAlaArgAspLeuMet 589
QY      2517 -----GTGCAAGAAGAATCCGCTTTTAAATCAACCCT---GAGTTGATTGTT 2561
Db      590 LeuMetSerHisLeuGlnAspAsnIleGlnHisAlaAspProProValGlnIleLeuTyr 609
QY      2562 TCACGGCTCTTCACTGAGGTG-----CTGGAT 2588
Db      610 AsnArgThrMetValGlnLeuGlyIleCysAlaPheArgGlnGlyLeuThrLysAspAla 629
QY      2589 CACAATGAATGCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCC 2648
Db      630 HisAsnAlaLeuLeuAspIleGlnSerSerGlyArgAlaLysGluLeuGlyGlnGly 649
QY      2649 GAAACATACCGCTGGCCCGGAGCCATGAAGACAGGCGAAAGGACGACGTCGACAGG 2708
Db      650 LeuLeuLeuArgSerLeuGlnGluArgAsnGlnGlnGluLysValGluArgArg 669
QY      2709 CGCAGCACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCACTTT 2768
Db      670 Gln----- 670
QY      2769 GTGCCCGAGAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTCTTCTGCTGGA 2828
```

```
Db      671 ValProPheHisLeuHisIleAsnLeuGluLeuLeuGluCysValTyrLeu----- 687
QY      2829 GTTTCCTCAGAGAACCTGTCTGCTGCTGCTACAGAAACTGGGGCTGTGGTGCCTTTGGG 2888
Db      688 -----ValSerAlaMetLeuLeuGluIleProTyrMetAlaAlaHisGlu 702
QY      2889 GGTGATGCTAGACTAAAGCCCTTAATACAGATCCTG-----GCAGCTGCTGTAGCTGAG 2942
Db      703 SerAspAlaArgArgMetIleSerLysGlnPheHisHisGlnLeuArgValGlyGlu 722
QY      2943 CGAGACGCTGCTTTATTTTCACCTTTGGGACTCAGAACTGATGAGA----- 2987
Db      723 ArgGlnProLeu-----LeuGlyProProGluSerMetArgGluHisValAla 739
QY      2988 -----GACATTACAGCATGTCATACATTCCTCACTGAGAGG 3023
Db      740 AlaSerLysAlaMetLysMetGlyAspTrpLysThrCysHisSerPheIleAsnGlu 759
QY      3024 AAACCTGACTGTTGGAGAAAGTATATAAGCTG----- 3053
Db      760 LysMetAsn---GlyLysValTrpAspLeuPheProGluAlaAspLysValArgThrMet 778
QY      3054 CTGCTACGATATTACAATGAAGATGCAGAAACTGCTCCACCCCGGACCAGACATCAAG 3113
Db      779 LeuValArgLysIleGlnGluGlu-----SerLeuArg 789
QY      3114 CTTTATCCATTCATATACCATGCAGTT 3140
Db      790 ThrTyrLeuPheThrTyrSerSerVal 798

RESULT 10
PVDA PLAKN          STANDARD;          PRT; 1073 AA.
ID PVDA PLAKN
AC P22545;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Duffy receptor, alpha form precursor (Erythrocyte binding protein).
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92357776; PubMed=1496004;
RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
RT "A family of erythrocyte binding proteins of malaria parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
RN [2]
RP SEQUENCE OF 296-1073 FROM N.A.
RX MEDLINE=91004213; PubMed=2170017;
RA Adams J.H., Hudson D.E., Torii M., Ward G.E., Wellem T.E.,
RA Aikawa M., Miller L.H.;
RT "The Duffy receptor family of Plasmodium knowlesi is located within
RT the micronemes of invasive malaria merozoites.";
RL Cell 63:141-153(1990).
CC -I- FUNCTION: Binds to the human erythrocytes Duffy blood group
CC determinant.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M90466; AAA29602.1; -.
CC EMBL; M68517; AAA29590.1; -.
CC EMBL; M68518; AAA29591.1; -.
CC PIR; A35970; A35970.
CC InterPro; IPR008602; Duffy_binding.
DR
```





SG1\_BOVIN  
ID SG1\_BOVIN STANDARD; PRT; 646 AA.  
AC P23389; O02707;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Secretogranin I precursor (SGI) (Chromogranin B) (CGB) [Contains: GAWK  
DE peptide; Secretolytin].  
GN CHGB.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adrenal chromaffin;  
RX MEDLINE=91223091; PubMed=2025642;  
RA Bauer J.W., Fischer-Colbrie R.;  
RT "Primary structure of bovine chromogranin B deduced from cDNA  
RL Biochim. Biophys. Acta 1089:124-126(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adrenal medulla;  
RX MEDLINE=97282588; PubMed=9136897;  
RA Yoo S.H., Kang Y.K.;  
RT "Identification of the secretory vesicle membrane binding region of  
RL chromogranin B";  
RN FEBS Lett. 406:259-262(1997).  
RN [3]  
RP SEQUENCE OF 21-646 FROM N.A.  
RC TISSUE=Adrenal medulla;  
RA Grandy D.K., Leduc R., Makam H., Flanagan T., Diliberto E.J.,  
RA Thomas G., Civelli O., Viveros O.H.;  
RL Submitted (OCT-1990) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 634-646.  
RC TISSUE=Adrenal chromaffin;  
RX MEDLINE=95262699; PubMed=7744058;  
RA Strub J.-M., Garcia-Sablone P., Lonning K., Taupenot L., Hubert P.,  
RA van Dorsselaer A., Aunis D., Metz-Boutigue M.-H.;  
RT "Processing of chromogranin B in bovine adrenal medulla.  
RT Identification of secretolytin, the endogenous C-terminal fragment of  
RT residues 614-626 with antibacterial activity.";  
RL Eur. J. Biochem. 229:356-368(1995).  
RN [5]  
RP CHARACTERIZATION OF SECRETOLYTIN.  
RX MEDLINE=96184581; PubMed=8603705;  
RA Strub J.M., Hubert P., Nullans G., Aunis D., Metz-Boutigue M.-H.;  
RT "Antibacterial activity of secretolytin, a chromogranin B-derived  
RT peptide (614-626), is correlated with peptide structure.";  
RL FEBS Lett. 379:273-278(1996).  
CC -!- FUNCTION: Secretogranin I is a neuroendocrine secretory granule  
CC protein, which may be the precursor for other biologically active  
CC peptides. The 16 pairs of basic AA distributed throughout its  
CC sequence may be used as proteolytic cleavage sites.  
CC -!- FUNCTION: Secretolytin has antibacterial activity.  
CC -!- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory  
CC granules.  
CC -!- PTM: O-glycosylated (Probable).  
CC -!- SIMILARITY: Belongs to the chromogranin / secretogranin protein  
CC family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X55027; CAA38846.1; -.

DR EMBL; U88551; AAC48720.1; -.  
DR EMBL; X55489; CAA39109.1; -.  
DR PIR; S15901; S15901.  
DR InterPro; IPR001819; Chromogranin\_AB.  
DR InterPro; IPR001990; Granin.  
DR Pfam; PF01271; Granin; 1.  
DR PRINTS; PR00659; CHROMOGRANIN.  
DR PROSITE; PS00422; GRANINS\_1; 1.  
DR PROSITE; PS00423; GRANINS\_2; 1.  
KW Sulfation; Cleavage on pair of basic residues; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 646 SECRETOGRANIN I.  
FT PEPTIDE 418 484 GAWK PEPTIDE.  
FT PEPTIDE 634 646 SECRETOLYTIN.  
FT DISULFID 36 57 BY SIMILARITY.  
FT MOD\_RES 158 158 SULFATION (POTENTIAL).  
FT MOD\_RES 315 315 SULFATION (BY SIMILARITY).  
FT CONFLICT 64 64 N -> S (IN REF. 1).  
FT CONFLICT 70 70 N -> D (IN REF. 2).  
FT CONFLICT 93 98 SEAPGL -> FRSPRAS (IN REF. 3).  
FT CONFLICT 181 181 T -> M (IN REF. 2).  
FT CONFLICT 261 261 H -> R (IN REF. 2).  
FT CONFLICT 386 386 P -> R (IN REF. 2).  
FT CONFLICT 481 481 H -> L (IN REF. 3).  
FT CONFLICT 597 597 M -> V (IN REF. 2).  
SQ SEQUENCE 646 AA; 73339 MW; 420DB1178FD9E415 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.0111 Length: 646  
Score: 164.50 Matches: 166  
Percent Similarity: 31.92% Conservatives: 90  
Best Local Similarity: 20.70% Mismatches: 269  
Query Match: 2.27% Indels: 277  
DB: 1 Gaps: 41  
  
US-09-302-812-1 (1-4070) x SG1\_BOVIN (1-646)  
QY 183 TCCGTCCTCCAGGGTTAGTGAATGAGGCTCTACGCCGGGCTGCCCGGAGACTCAGT 242  
DB 49 SerAlaProPro-----IleThrProGluCysArgGlnValLeuLys 62  
QY 243 GCTGCGGGTCCCAGCATGATGATGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGC 302  
DB 63 LysAsnGlyLysGluLeuLysAsnGluGluLysSerGlu-----AsnGluAsnThrArg 80  
QY 303 TGGGACGCGCTGCAACTTCTCCGCGCGCGCTCGGACGCCCGGAGCTTCCCGCGCAGG 362  
DB 81 PheGluValArgLeuLeuArgAspProAla-----AspThrSerGluAlaProGlyLeu 98  
QY 363 CAGAGCGCGTCTCTCGATTCC-----AAGGACGCTCCGGTGCAGTTCAGGGTCCCGCGC 416  
DB 99 SerSerArgGlu--AspSerGlyGluGlyAspAlaGlnValProThrValAlaAspThr 117  
QY 417 TCCTCGTCAGGCTGCGCCCTGGCGCGGGCGGACAGCACCGAGGACGCCACCTCTCTT 476  
DB 118 GluSerGlyGlyHisSerArgGluArgAlaGlyGluProProGlySerGlnValAla--- 136  
QY 477 GTTTTCAAACAGAGACTATACCAGTTGGATGGACACTAAAGCAATCAAGACAGTTGAA 536  
DB 137 -----LysGlu 138  
QY 537 TCAGAAAGTTGCATAGTAAAGAAACAAACAATACAAGAGAGAAATCCATGATGATTCT 596  
DB 139 AlaLysThrArgTyrSerLysSerGluGlyGlnAsnArgGluGluMetValLys--- 157  
QY 597 GTACAAAAGATAAATCTTTTATCAATACATGAAATAATAGAAAATGTTTCTCAGCTA 656  
DB 158 -----TyrGln---LysArgGluArgGlyGluValGlySerGluGlu 170  
QY 657 GGTTTTGATAACTCACCAGTTGAAAAGGTGACAGTATTTGAAGCAGCATCAGACTGCG 716  
DB 171 ArgLeuSerGluGlyProGlyLysAlaGlnThrAlaPheLeuAsnGln----- 186



|    |      |  |      |                |  |  |          |
|----|------|--|------|----------------|--|--|----------|
| QY | 717  | GCTATGTGTAAGTGGCAGAATGAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCT   | 776  | Db             | 421  | GluserglnargaspLysAlaArgArgLeuProGlyGluLeuArgAsnTyrLeuAsp        | 440      |
| Db | 187  | -----ArgAsnGlnThrProAlaLysLysGluGluLeuValSerArg---             | 200  | QY             | 1827   | GTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACATTCAGACTGCA       | 1886     |
| QY | 777  | CCAGCGGTAACTCTGGTACCAGAGCAGTTTCACTAATGTCGATCAGTCGTCCCCA        | 836  | Db             | 441  | TyrGlyGluGluLysGlyGluAlaAlaArgGlyLysTrpGln-----                  | 455      |
| Db | 201  | -----TyrAspThrGlnSerAlaArgGlyLeuGluLysSerHisSer                | 214  | QY             | 1887   | CTTCTCAACAGGCTCACTCGGCCCCAG---AACCTGAAGGATCTATTCTGAAGTACAAT      | 1943     |
| QY | 837  | AAGGATGATCAGCTGACACAAATAGTGAGGAGAGTAGATAATCAGCAGTTTGTGACA      | 896  | Db             | 456  | -----ProGlnGlyAspProArgAspAla-----                               | 463      |
| Db | 215  | ArgGluArgSerGlnGluSerGlyGluGluThrLysSerGlnGluAsnTrp-----       | 232  | QY             | 1944   | GTGGCATATTTAAGAAATGGGACTTTACAGCTTTTGATTGATTCTGGGATAAGGTACTA      | 2003     |
| QY | 897  | CATGTAAGCTTGGCAATGCAAGCAGACGATGGAAGATGAACAGGGCAGAGCCAGA        | 956  | Db             | 464  | -----AspGluAsnArg  | 467      |
| Db | 233  | -----ProGlnGluLeuGlnArgHisProGluGlyGlnGluAla---                | 245  | QY             | 2004   | GAAGAAGCAGAAAGCTCAACACTTGTATCAGTCCATCTTGCTGATATGGTGAATAATTGCA    | 2063     |
| QY | 957  | AGCCACCAGAAGTGTGGCAAGGCTTGCCATCTGCAGAACGCTGTGCAGGGTGTGACGAG    | 1016 | Db             | 468  | GlulGluAlaArgLeuArgGly-----                                      | 474      |
| Db | 246  | -----ProGlyGluSer-----   | 249  | QY             | 2064   | CTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACCTCCTGAAACAGAGATGAATCAT     | 2123     |
| QY | 1017 | GAGGAGACAGACGTGTCTCCGAGAGCCCCCTTGTGCGACACTGGCTCTGAGGATGTTGGT   | 1076 | Db             | 475  | -----LysGlnTyrAlaProHis  | 480      |
| Db | 250  | ---GluGluAspAla-----SerProGluValAsp-----                       | 258  | QY             | 2124   | TCCATCACAATGTACAGAAACAGATTGGCCAGTCTTTTAGCTAATGCTTTCTTC           | 2177     |
| QY | 1077 | ACTGGACTGAAAATGCCAACAGATTGAATAGACAAAGAAAGTAGTCTAGGAAATTTCTCT   | 1136 | Db             | 481  | HisIleThr-----GluLysArgLeuGlyGluLeuLeu---AsnProPheTyrAspPro      | 497      |
| Db | 259  | -----LysArgHisSerArgProArgHisHisGlyArgSerArg                   | 272  | QY             | 2178   | -----TGCACGTTTCCACGACGC-----                                     | 2195     |
| QY | 1137 | CCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGATAATTCAAAAATAGT    | 1196 | Db             | 498  | SerGlnTrpLysSerSerArgPheGluArgLysAspProMetAspSerPheLeuGlu        | 517      |
| Db | 273  | ProAspArgSerSerGlnGluGlyAsnProProLeu-----                      | 284  | QY             | 2196   | -----AATGCCAAGATGAAATCAGAGTATCCAGATTATCCAGATTATTAATTCTTC         | 2243     |
| QY | 1197 | TGTCAGGATTGAGAGCAGATGAAGAGACAAAGTCCAGGTTTGTATGAACAGGAAGTAGC    | 1256 | Db             | 518  | GlyGluGluGluAsnGlyLeuThrLeuAsnGluLysAsnPheProGluTyrAsnTyr        | 537      |
| Db | 285  | -----GluGluGluSerHisValGly-----ThrGlyAsn                       | 294  | QY             | 2244   | AAT-----CGGTTGTTTGAA-----GGACGTTTCATCAAGGAAA                     | 2276     |
| QY | 1257 | AGTTCTGCTCAACAGCAAAATAAACCCTTCAAGGTTCCAAACCAGAGAAAGTGCACACTGAG | 1316 | Db             | 538  | AspTrpTrpGluLysLysProPheGluGluAspValAsnTrpGlyTyrGluLysArgAsn     | 557      |
| Db | 295  | SerAspGluGluLysAlaArgHisProAlaHisPheArgAlaLeuGluGluGlyAlaGlu   | 314  | QY             | 2277   | CCAGAGAAAGCTTAAACCGCTCTTCTGCTACTTTAGAGAGTCCACAGAGAAAACCCACT      | 2336     |
| QY | 1317 | TTGAGGAACGGTCTCTCTGCTAAGGGAGGTGAGATTGATTACATTTCCAATTTGAAGGA    | 1376 | Db             | 558  | ProValProLysLeuAspLeuLysArgGlnTyrAspArgValAlaGluLeuAspGln---     | 576      |
| Db | 315  | Tyr-----GlyGluGluValArgArgHisSerAla-----                       | 324  | QY             | 2337   | GGGTTGGTGACATTCAACAGACAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAA     | 2396     |
| QY | 1377 | GGAGAGAGTCGAGCTGGAATGAATGATGTGAATGCCAAACGACCTCGAAGTACTTCTAGC   | 1436 | Db             | 577  | ---LeuLeuHisTyrArgLysLysSerAlaGlu---PheProAspPheTyrAspSerGlu     | 594      |
| Db | 325  | -----AlaGlnAlaProGlyAspLeuGlnGly                               | 333  | QY             | 2397   | AAACTCCTGACTCGACTGCTGCTACTTACGAGGTACGAAAGTACCATAGAAGAAACGGCCAGGC | 2456     |
| QY | 1437 | CTGAATGTAGAGTCAGAAAATTTCTAAGCAACATGGGAGAAAGGATTTCTAAATCAACAT   | 1496 | Db             | 595  | GluGlnMetSerProGlnHis---ThrAlaGluAsnGluGluLysAlaGlyGlnGly        | 613      |
| Db | 334  | AlaArgPheGlyGlyArgGlyArgGlyGluHis-----                         | 344  | QY             | 2457   | ATGCTA   | 2462     |
| QY | 1497 | CATTTTCATGAGTGTCCCAAGCAGAGAGGACAAAGAAAGAAACAATGTGAATGAAACAT    | 1556 | Db             | 614  | ValLeu   | 615      |
| Db | 345  | GlnAlaLeuArgArgProSerGluGluSer-----LeuGluGlnGluAsnLysArgHis    | 362  | RESULT 12      |  |  |          |
| QY | 1557 | CAAAGAACAGAAAGGAGATCCCTAAATACATTCACCTCACCTTCTCCAGATAAGAAA      | 1616 | PVDG_PLAKN     |  |  |          |
| Db | 363  | -----GlyLeuSerProAspLeuAsn                                     | 369  | ID_PVDG_PLAKN  | STANDARD;  | PRT;   | 1070 AA. |
| QY | 1617 | TGGCTTGGAACTCCTATTGAGGAGATG-----AGGAGAATGCCAAGGTGTGGATCCGG     | 1670 | AC_P50494;     |  |  |          |
| Db | 370  | MetAlaGlnGlyTyrSerGluGluSerGluGluGlyProAlaProGly-----          | 387  | DT_01-OCT-1996 | (Rel. 34, Created)   |  |          |
| QY | 1671 | CTGCCTCCCTTGAGA-----CCATCTGCCAATCACACAGTACTATTCCGGTA           | 1718 | DT_01-OCT-1996 | (Rel. 34, Last sequence update)                                      |  |          |
| Db | 388  | ---ProSerTyrArgAlaArgGlyGlyGluAlaAlaTyrSerThrLeuGlyGlnThr      | 406  | DT_01-OCT-1996 | (Rel. 34, Last annotation update)                                    |  |          |
| QY | 1719 | GATCTTTTGGCA---ATAGGAGAAAGTTCTTAAACCTTTCCCAACACATTTTAAA        | 1769 | DE             | Duffy receptor, gamma form precursor (Erythrocyte binding protein).  |  |          |
| Db | 407  | AspGluLysArgPheLeuGlyGlu-----ThrHisHisArgValGln                | 420  | OS             | Plasmodium knowlesi.   |  |          |
| QY | 1770 | ---GATTTGTGGGACAAACAGCATGTTAAGATGCCTTGTTCAGAAACAAACTTGACCT     | 1826 | OC             | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.         |  |          |
|    |      |  |      | OX             | NCBI_TaxID=5850;   |  |          |
|    |      |  |      | RN             | [1]  |  |          |
|    |      |  |      | RP             | SEQUENCE FROM N.A.   |  |          |
|    |      |  |      | RX             | MEDLINE=92357776; PubMed=1496004;                                    |  |          |
|    |      |  |      | RA             | Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.; |  |          |
|    |      |  |      | RT             | "A family of erythrocyte binding proteins of malaria parasites."     |  |          |
|    |      |  |      | RL             | Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).                    |  |          |
|    |      |  |      | CC             | -!- FUNCTION: Binds to the human erythrocytes Duffy blood group      |  |          |



Db 923 GluLysIleIleLeuMetSerGluValAsnLysCysAsnAsnArgThrSerLeuLysTyr 942  
::: ::: :::  
::: ||| :::  
::: ||| :::

QY 1557 CAAAGAAC 1565  
|||  
Db 943 CysAsnThr 945

RESULT 13  
BRD4 HUMAN  
ID BRD4 HUMAN STANDARD; PRT; 1362 AA.  
AC O6085; Q96PD3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Bromodomain-containing protein 4 (HUNK1 protein).  
GN BRD4 OR HUNK1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=21590020; PubMed=11733348;  
RA French C.A., Miyoshi I., Aster J.C., Kubonishi I., Kroll T.G.,  
RA Dal Cin P., Vargas S.O., Perez-Atayde A.R., Fletcher J.A.;  
RT "BRD4 bromodomain gene rearrangement in aggressive carcinoma with  
translocation t(15;19).";  
RL Am. J. Pathol. 159:1987-1992(2001).  
[2]  
SEQUENCE OF 1-722 FROM N.A.  
RP TISSUE=Placenta;  
RA Weber B.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- SIMILARITY: Contains 2 bromodomains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF386649; AAL26987.1; --  
CC EMBL; Y12059; CAA72780.1; --  
CC HSSP; Q92831; 1B91.  
CC Genew; HGNC:13575; BRD4.  
CC InterPro; IPR001487; Bromodomain.  
CC Pfam; PF00439; bromodomain; 2.  
CC PRINTS; PR00503; BROMODOMAIN.  
CC SMART; SM00297; BROMO; 2.  
CC PROSITE; PS00633; BROMODOMAIN\_1; 1.  
CC PROSITE; PS50014; BROMODOMAIN\_2; 2.  
KW Bromodomain; Repeat; Nuclear protein.  
FT DOMAIN 75 147 BROMODOMAIN 1.  
FT DOMAIN 368 440 BROMODOMAIN 2.  
FT DOMAIN 535 594 LYS-RICH.  
FT DOMAIN 692 717 SER-RICH.  
FT DOMAIN 703 714 POLY-SER.  
FT DOMAIN 738 743 POLY-HIS.  
FT DOMAIN 757 761 POLY-PRO.  
FT DOMAIN 764 770 POLY-PRO.  
FT DOMAIN 771 775 POLY-GLN.  
FT DOMAIN 776 783 POLY-PRO.  
FT DOMAIN 954 964 POLY-PRO.  
FT DOMAIN 974 986 POLY-PRO.  
FT DOMAIN 1011 1014 POLY-PRO.  
FT DOMAIN 1028 1033 POLY-PRO.  
FT DOMAIN 1283 1300 POLY-GLN.  
FT DOMAIN 1301 1308 POLY-ALA.  
FT DOMAIN 1335 1338 POLY-ARG.  
FT CONFLICT 720 721 EM -> GP (IN REF. 2).

SQ SEQUENCE 1362 AA; 152219 MW; D52EFE1CF9960907 CRC64;  
Alignment Scores: Length: 1362  
Pred. No.: 1362  
Score: 163.00 Matches: 119  
Percent Similarity: 37.12% Conservative: 100  
Best Local Similarity: 20.17% Mismatches: 228  
Query Match: 2.25% Indels: 144  
DB: 1 Gaps: 26  
US-09-302-812-1 (1-4070) x BRD4\_HUMAN (1-1362)  
QY 156 CCGGAGGACCGAGGAGGCGCGAGTCCGTCCTCCAGGTTAGTGAATGAGGCTCT 215  
Db 261 ProAlaProAlaProGlnProValGlnSerHisProProIleAla-----Ala 277  
QY 216 AGCCCC-----GGGCTGGCCCGGAGACTCAGTGTGCGGGTCCC 254  
Db 278 ThrProGlnProValLysThrLysLysGlyValLysArgLysAlaAspThrThrPro 297  
QY 255 AGCATAGTTCGGGCGCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGC----- 302  
Db 298 Thr---ThrIleAspProIleHisGlnProProSerLeuProProGluProLysThr 316  
QY 303 -----TGGGACGCGCTGCAACTTCTCCGCGCGCGCTCGGACGCGCGGAGCTTCCC 356  
Db 317 LysLeuGlnArgArgGluSerSerArgProValLysProProLysLysAspValPro 336  
QY 357 GGCAGGCAGAGCGCGTCTCGATTCCAGGACGCTCCGGTGCAGTTTCAAGGTCGCCG 416  
Db 337 AspSerGlnGlnHisProAlaProGluLysSerLysValSerGluGlnLeuLysCys 356  
QY 417 TCCTCGTCAGGCTGCGCCCTGGCGCGGCGGAGCAGCACCGAGGCGGCGACCTCTCT 476  
Db 357 CysSer-----GlyIleLeuLysGluMet 364  
QY 477 GTTTTCAACAGAGACTATAACAGTTGG----- 506  
Db 365 PheAlaLysLysHisAlaAlaTyrAlaTrpProPheTyrLysProValAspValGluAla 384  
QY 507 -----ATGGACACTAAAGGAATCAAG 527  
Db 385 LeuGlyLeuHisAspTyrCysAspIleIleLysHisProMetAspMetSerThrIleLys 404  
QY 528 ACAGTTGAATCAGAAAGTTTGCATAGTAAAGAAACAACAATACAGAGAAATCCATG 587  
Db 405 Ser-----LysLeuGluAlaArgGluTyrArgAspAlaGlnGluPheGlyAla 420  
QY 588 ATGAGTTCTGTACAAAAGATAAATTTTATCAACATAAC----- 626  
Db 421 AspValArgLeuMetPheSerAsnCysTyrLysTyrAsnProProAspHisGluValVal 440  
QY 627 -----ATGGAAAATAGAAAATGTTTCTCAGTAGTGTGTTTGTATTAAGTCAACAG 680  
Db 441 AlaMetAlaArgLysLeuGlnAspValPheGluMetArgPheAlaLysMetProAspGlu 460  
QY 681 AAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAACTGCGGAGATGAA 740  
Db 461 ProGluGluProValValAlaValSerSerProAlaVal----- 473  
QY 741 GGGCCACACTCAGAACGGCTTTTGGAAAAGTGAACCTCCAGCGGTAACTCTGGTACCAGAG 800  
Db 474 -----ProProThrLysValValAlaProPro 483  
QY 801 CAGTTTCAGTAAATGTCGATCAGTCGTCCTCCCAAGGATGATCAGTGCACACAAAT 860  
Db 484 SerSerSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 502  
QY 861 AGTGAGGAGAGTAGA-----GATAATCAGCAGTTTGTGACACAT 899  
Db 503 SerGluGluGluArgAlaGlnArgLeuAlaGluLeuGlnGluLeuLysAlaValHis 522  
QY 900 GTAAAGCTTGGCAATGCAAGACGACGATGGAAGATGAACAGGGCAGAGAGCCAGAAGC 959



Db 523 GluGlnLeuAlaLeuSerGlnProGlnGlnAenLysProLysLysLysGluLysAsp 542  
QY 960 CACCAGAAGTGTGGCAAGCTTGCCATCTGCAGAACCTGTGCAGGGTGTGCAGCAGGAG 1019  
Db 543 LysLysGluLysLysLysGluLysHisLysArgLys-----GluGluValGluGluAsn 560  
QY 1020 GAGACAGACGTGGTGTCCGAGAGCCCTTGTCCGACACTGGCTCTGAGGATGTGGTACT 1079  
Db 561 LysLysSerLysAlaLysGluProProLysLysThrLysLysAsnAsn----- 577  
QY 1080 GGACTGAAAAATGCCAACAGATTGAATAGACAAGAACTAGTCTAGGAAATTCCTCTCCA 1139  
Db 578 ---SerSerAsnSerValSerLysLysGluProAlaProMetLysSerLysProPro 596  
QY 1140 TTTGAGAAAGAAAGTGAACCTGAG-----TCACCAATGGATGTAGATAATTCCAAA 1190  
Db 597 ProThrTyrGluGluLysArgGlnLeuSerLeuAspLysCysLysProMet----- 610  
QY 1191 AATAGTTGTCAGGATTTCAGAACAGCATGAAGAGACAAAGTCCAGGTTTGTG----- 1239  
Db 611 ---SerTyrGluGluLysArgGlnLeuSerLeuAspLysLeuProGlyGluLys 629  
QY 1240 -ATGAACAGGAAGATAGCAGTTCTGCTCAACAGCAATAAACCTTCAAGGTTCCAAACCA 1298  
Db 629 sLeuGlyArgValHisIleIleGlnSerArgGluProSerLeuLysAsnSerAsnPr 649  
QY 1299 AGAGAAGCTG-----ACACTGAGTTGAGGAGCGGCTCTCTGCTAAGGGA 1343  
Db 649 oAspGluIleGluIleAspPheGluThrLeuLysProSerThr-----LeuArgG 666  
QY 1344 GGTGAGATTGATTACATTTCGAATTGAGAGGAGAGAGT-----CGAGCTGGAATGAA 1399  
Db 666 uLeuGluArgTyrValThrSerCysLeuArgLysLysArgLysProGlnAlaGluLysVa 686  
QY 1400 TGATGTG-----AATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTG 1450  
Db 686 lAspValIleAlaGlySerSerLysMetLysGlyPheSerSerSer-----GluSe 703  
QY 1451 CAGAAATTCTAAGCAACATGGGAGAAAGGATTCTAAATACACAGATCATTTTCATGAGAGT 1510  
Db 703 rGluSerSerSerGluSerSerSerSerAspSerGluAspSerGluThrGluMet---Al 722  
QY 1511 GCCCAACAGCAGGACAAA-----AGAAAAGAACAAATGTGAAATGAAACAT----- 1556  
Db 722 aProLysSerLysLysLysGlyHisProGlyArgGluGlnLysLysHisHisHisHi 742  
QY 1557 -----CAAAGAACAGAAAGGAAGATCCCTAAATACATTCACCTCACCTTTC 1603  
Db 742 sHisGlnGlnMetGlnGlnAlaProAlaProValProGlnGlnProPro-----Pr 760  
QY 1604 TCCAGATAAGAAATGGCTTGGAATCTTATTGAGGAGATGAGGAGAATGCCAAGGTGG 1663  
Db 760 oProGlnGlnProProProProProProProProGlnGlnGlnGlnProProPro----- 778  
QY 1664 GATCCGGCTGCTCCCTTGAGACCATCT 1691  
Db 779 -----ProProProProSer 784

RESULT 14  
DMP1\_HUMAN  
ID DMP1\_HUMAN STANDARD; PRT; 513 AA.  
AC Q13316; O43265;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix  
protein-1) (DMP-1).  
GN DMP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

[1]  
SEQUENCE FROM N.A. (ISOFORM 1).  
TISSUE=Molar;  
MEDLINE=97321043; PubMed=9177774;  
RA Hirst K.L., Simmons D., Feng J., Aplin H., Dixon M.J., McDougall M.;  
RT "Elucidation of the sequence and the genomic organization of the human  
dentin matrix acidic phosphoprotein 1 (DMP1) gene: exclusion of the  
RT locus from a causative role in the pathogenesis of dentinogenesis  
imperfecta type II.";  
RL Genomics 42:38-45(1997).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 2).  
McDougall M., Juan X., Simmons D., Feng J.;  
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 462-513 FROM N.A., AND GENE MAPPING.  
MEDLINE=96163890; PubMed=8586437;  
RA Aplin H.M., Hirst K.L., Crosby A.H., Dixon M.J.;  
RT "Mapping of the human dentin matrix acidic phosphoprotein gene (DMP1)  
to the dentinogenesis imperfecta type II critical region at chromosome  
4q21.";  
RL Genomics 30:347-349(1995).  
CC -!- FUNCTION: Could be involved in the induction of mineralization of  
extracellular matrix.  
CC -!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=1;  
IsoId=Q13316-1; Sequence=Displayed;  
Name=2;  
IsoId=Q13316-2; Sequence=VSP\_004191;  
CC -!- TISSUE SPECIFICITY: Expressed in tooth particularly in  
odontoblast, ameloblast and cementoblast.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U89012; AAC51332.1; -.  
EMBL; U34037; AAA97602.1; -.  
EMBL; U65378; AAB87728.1; -.  
Genew; HGNC:2932; DMP1.  
MIM; 600980; -.  
GO; GO:0005578; C:extracellular matrix; TAS.  
GO; GO:0005509; F:calcium ion binding; TAS.  
GO; GO:0005178; F:integrin binding; TAS.  
GO; GO:0001503; P:ossification; TAS.  
KW Biomaterialization; Extracellular matrix; Signal; Alternative splicing.  
FT SIGNAL 1 16  
FT CHAIN 17 513  
FT CARBOHYD 25 25  
FT CARBOHYD 285 285  
FT CARBOHYD 324 324  
FT CARBOHYD 345 345  
FT CARBOHYD 351 351  
FT CARBOHYD 413 413  
FT CARBOHYD 426 426  
FT CARBOHYD 467 467  
FT SITE 364 61  
FT VARSPLIC 46 61  
FT DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CELL ATTACHMENT SITE (POTENTIAL).  
FT Missing (in isoform 2).  
FT /FTId=VSP\_004191.  
FT S -> C (IN REF. 2).  
FT CONFLICT 69 69  
SQ SEQUENCE 513 AA; 55782 MW; 2C1FDE319A5D106F CRC64;  
  
Alignment Scores:  
Pred. No.: 0.0149 Length: 513  
Score: 162.00 Matches: 101  
Percent Similarity: 35.73% Conservative: 78  
Best Local Similarity: 20.16% Mismatches: 194  
Query Match: 2.24% Indels: 128



```
DB: 1 Gaps: 20
US-09-302-812-1 (1-4070) x DMP1_HUMAN (1-513)
QY 114 GCAGGAGCTGCAGAAGCAGTACGCGGAGAGGGGGCATGGTCCGGGAGGCACCGAGGAG 173
Db 88 AlaGlyGlyPheSerArgSerThrGlyLysGlyGlyAspAspLysAspAspGluAsp 107
QY 174 GGGGCGCAGTCCGTCCTCCAGGGTTAGTAATGAGGCTCTACGCCCGGGGTGGCCCGG 233
Db 108 -----AspSerGlyAspAspThrPheGlyAspAspAspSerGlyProGlyProLysAsp 125
QY 234 AGACTCAGTGTGCGGGTCCAGCATGAGTGGGGCCCGGCTGTGAGCCCTGCACCAAG 293
Db 126 ArgGlnGluGlyGlyAsnSerArgLeuGlySerAspGluAspSerAspThrIleGln 145
QY 294 CGACCCCGTGGGCGCGCTGCAACTTCTCCGCGCGCGCTCGGACGCCCGGAGCTTC 353
Db 146 -----AlaSerGluGluSerAlaProGlnGlyGlnAspSerAlaGlnAsp 160
QY 354 CCCGGCAGGCAGAGCGGCTCCTCGATTCCAGGACGCTCCGGTGCAGTTCAGGGTCCCG 413
Db 161 ThrThrSerGluSerArgGluLeuAspAsnGluAsp-----ArgValAsp 175
QY 414 CCGTCCTCGTCAGGCTGCGCCCTGGCGCGGGGACAGCACCGGAGCGCCACCTCT 473
Db 176 SerLysProGluGly-----GlyAspSerThrGln 185
QY 474 CTTGTTTCAAAACAGAGACATATAACAGTTGG----- 506
Db 186 -----GluSerGluSerGluGluHisTyrValGlyGlySerAspGlyGluSer 202
QY 507 -----ATGCACACTAAAGGAATCAAGACAGATTGAATCAGAAAGT 545
Db 203 SerHisGlyAspGlySerGluLeuAspAspGluGlyMetGlnSerAspAspProGluSer 222
QY 546 TTGCATAGTAAGAAAACAACATACAAGAGAGAAATCCATGATGAGTTCTGTACAAAAA 605
Db 223 IleArgSerGluArgGlyAsnSerArgMetAsnSerAlaGlyMetLysSerLysGluSer 242
QY 606 ---GATAACTTTTATCAACATAACATGGAATA-----TTAGAAAT 644
Db 243 GlyGluAsnSerGluGlnAlaAsnThrGlnAspSerGlyGlySerGlnLeuLeuGluHis 262
QY 645 GTTTCCTCAGTAGTTTGTATAGTCACCGAGTTGAAAAGGTACACAGTATTGAGGCAG 704
Db 263 ProSerArgLysIlePheArgLysSerArgIleSerGlu----- 275
QY 705 CATCAGACTGGGCTATGTGTAAGTGGCAGAAATGAAGGGCCACACTCAGAACGGCTTTTG 764
Db 276 -----GluAspAspArgSerGlu-----Leu 282
QY 765 GAAAGTGAACCTCCAGCGGTAACTCTGTACAGAGCAGTTCAGTAATGCTAAT----- 818
Db 283 AspAspAsnAsnThrMetGluGluValLysSerAspSerThrGluAsnSerAsnSerArg 302
QY 819 ---GTCGATCAGTCGTCCCCCAAGGATGATCAGAGTGCACACAAATAGTGAGGAGTAGA 875
Db 303 AspThrGlyLeuSerGlnProArgArgAspSerLysGlyAspSerGlnGluAspSerLys 322
QY 876 GATAATCAGCAGTTTGTGACACATGTAAAGCTTGCGAATGCAAGCAGACGATGGAAGAT 935
Db 323 GluAsn-----LeuSerGlnGluGlu 329
QY 936 GAACAGGGCAGAGAAGCCAGAGCCACCAGAAGTGTGGCAAGGCTTGCCATCCTGCAGAA 995
Db 330 SerGlnAsnValAspGlyProSerSerGluSerSerGlnGluAlaAsnLeuSerSerGln 349
QY 996 GCCTGTGCAGGGTCTCAGCAGGAGGAG----- 1022
Db 350 GluAsnSerSerGluSerGlnGluValValSerGluSerArgGlyAspAsnProAsp 369
QY 1023 -----ACAGACGTGTGTCGAGAGCCCTTCTCGGACACTGGCTCTGAGGATGTTGGT 1076
```

```
Db 370 ProThrThrSerTyrValGluAspGlnGluAspSerAspSerSerGluGluAspSerSer 389
QY 1077 ACTGGACTGAAAAATGCCAACAGATTGAATAGACAAGAA-----AGTAGTCTAGGA 1127
Db 390 HisThrLeuSerHisSerLysSerGluSerArgGluGlnAlaAspSerGluSerSer 409
QY 1128 AATTCTCTCCATTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGATAATTCC 1187
Db 410 GluSerLeuAsnPheSerGluGluSer---ProGluSerProGluAspGluAsnSerSer 428
QY 1188 AAAAAAT-----AGTTGTCAGGATTCAAGACGAGATGAAGACAGCAAGTCCAGGTTT 1238
Db 429 SerGlnGluGlyLeuGlnSerHisSerSerSerAlaGluSerGlnSerGluGluSerHis 448
QY 1239 GATGAACAGAGAGATAGCAGTTCTGCTCAAAACAGCAAATAAACCTTCAAGGTTCCAAACCA 1298
Db 449 SerGluGluAspAspSerAspSerGlnAspSer-----SerArgSerLysGlu 464
QY 1299 AGAGAAGCTGCACACTGAGTTTGAGGAAGCGGTCTCTGCTAAGGGAGGTGAGATTCGATTA 1358
Db 465 AspSerAsnSerThrGlu---SerLysSerSerSerGluGluAspGlyGlnLeuLys--- 482
QY 1359 CATTTCCAAATTTGAAGGAGGAGAGAGTGGAGCTGCTGAATGAATGATGTGAATGCCAAACGA 1418
Db 483 AsnIleGluIleGlu----- 487
QY 1419 CCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTTCTAAGCAACATGGGAGAAAAG 1478
Db 488 -----SerArgLysLeuThrValAspAlaTyrHisAsnLysProIleGlyAspGln 504
QY 1479 GAT 1481
Db 505 Asp 505
RESULT 15
MY9B_HUMAN STANDARD; PRT; 2158 AA.
ID MY9B_HUMAN Q13459; O75314; Q9NUJ2; Q9UHN0;
AC Q13459; O75314; Q9NUJ2; Q9UHN0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin Ixb (Unconventional myosin-9b).
GN MYO9B OR MYR5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver, and Small intestine;
RX MEDLINE=97063843; PubMed=8907710;
RA Wirth J.A., Jensen K.A., Post P.L., Bement W.M., Mooseker M.S.;
RT "Human myosin-IXb, an unconventional myosin with a chimerin-like
RT rho/rac GTPase-activating protein domain in its tail.";
RL J. Cell Sci. 109:653-661(1996).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=20047919; PubMed=10580159;
RA Grewal P.K., Jones A.-M., Maconochie M., Lemmers R.J.F., Frants R.R.,
RA Hewitt J.E.;
RT "Cloning of the murine unconventional myosin gene Myo9b and
RT identification of alternative splicing.";
RL Gene 240:389-398(1999).
RN [3]
RP SEQUENCE OF 1940-2158 FROM N.A. (ISOFORM LONG).
RX MEDLINE=98158729; PubMed=9490638;
RA Post P.L., Bokoch G.M., Mooseker M.S.;
RT "Human myosin-IXb is a mechanotchemically active motor and a GAP for
RT rho.";
RL J. Cell Sci. 111:941-950(1998).
RN [4]
RP SEQUENCE OF 1828-2023 FROM N.A. (ISOFORM SHORT).
```



QY 519 GGAATCAAG-----ACAGTTGAATCAGAAAGTTGTCATAGTAAAGAA 560  
Db 1174 AlaLeuArgGluProSerArgValThrGlnGluGlnGlyValSerLeuLeuGluAsp 1193  
QY 561 AACAAACAATACAAAGAGAAGAAATCCATGATGAGTTCTGTACAAAAAGATAAACTTTTATCAA 620  
Db 1194 LysLysGluSerArgGluAspGluThrLeuLeuValValGluThrGlu----- 1209  
QY 621 CATAACATCGAAAAAATTAGAAAAATGTTTCTCAGCTAGGTTTGTGATAAGTCACCAGTTGAA 680  
Db 1210 -----AlaGluAsnThrSerGln-----LysGlnProThrGlu 1220  
QY 681 AAAGGTACACAGTATTTGAAGCAGCATCAGACTCGGGCTATGTGTAAAGTGGCAGAATGAA 740  
Db 1221 -----GlnProGlnAlaMetAlaValGlyLysValSerGluGlu 1233  
QY 741 GGGCCACACTCAGAACGGCTTTTGGAAAGT----- 770  
Db 1234 -----ThrGluLysThrLeuProSerGlySerProArgProGlyGlnLeuGluArg 1250  
QY 771 -----GAACCTCCAGCGGTAACTCTGGTACCAGAG 800  
Db 1251 ProThrSerLeuAlaLeuAspSerArgValSerProProAlaProGlySerAlaProGlu 1270  
QY 801 CAGTTTCAGTAATGCTAATGTCGATCAGTCGTCCCAAGGATGATCAGAGTGACACAAAT 860  
Db 1271 ThrProGluAspLysSerLysProCysGlySerProArgValGlnGluLysProAspSer 1290  
QY 861 AGTGAGGAGAGTAGAGATAATCAGCAGATTTTGGACACATGTAAAGCTTGCGAATGCAAG 920  
Db 1291 ProGlyGlySerThrGlnIleGlnArgTyrLeuAspAlaGluArgLeuAlaSerAlaVal 1310  
QY 921 CAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAGC----- 959  
Db 1311 GluLeuTyrArgGlyLysLysLeuValAlaAlaSerProSerAlaMetLeuSerGln 1330  
QY 960 -----CACCAGAAAGTGTGGCAAGCTTGCCATCTCTCAGAAAGCC 998  
Db 1331 SerLeuAspLeuSerAspArgHisArgAlaThrGlyAlaAlaLeuThrProThrGlu--- 1349  
QY 999 TGTGCAGGTGTGCAGGAGGAGAGACAGACGTGTGTCGAGAGCCCTTGTTCGACACT 1058  
Db 1350 -----GluArgArgThrSerPheSerThrSerAspValSer----- 1361  
QY 1059 GGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAACAGATTGAATAGACAAAGAAAGT 1118  
Db 1362 -----LysLeuLeuPro 1365  
QY 1119 AGTCTAGGAAATTCCTCCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGAT--- 1175  
Db 1366 SerLeuAlaLysAlaGlnProAlaAlaGluThrThrAspGlyGluArgSerAlaLysLys 1385  
QY 1176 -----GTAGATAATTCCTCAAAAATAGTTGTTCAGGATTCAGAAAGCAGATGAAGAGACAAGT 1229  
Db 1386 ProAlaValGlnLysLysLysProGlyAspAlaSerSerLeuProAspAlaGlyLeuSer 1405  
QY 1230 CCAGGTTTGTATGAACAGGAGATAGCAGTTCTGCT-----CAA 1268  
Db 1406 ProGly-----SerGlnValAspSerLysSerThrPheLysArgLeuPheLeuHisLys 1423  
QY 1269 ACAGCAATAAAACCTTCAAGGTTTCAACCAAGAGAA----- 1304  
Db 1424 ThrLysAspLysLysTyrSerLeuGluGlyAlaGluGluLeuGluAsnAlaValSerGly 1443  
QY 1305 -----GCTGACACTGAGTTGAGGAAGCGGTCTCTCTCTAAGGGAGGTGAGATT 1352  
Db 1444 HisValValLeuGluAlaThrThrMetLysLysGlyLeuGluAlaProSerGlyGlnGln 1463  
QY 1353 CGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATGAATGATGTGAATGCC 1412  
Db 1464 HisArgHis-----AlaAlaGlyGluLysArg-----Thr 1473

QY 1413 AAACGACACTGGAAGTACTTCTAGCCTGAATGTAGAGTGACAGAAATTTAAGCAACATGGG 1472  
Db 1474 LysGluProGlyGly-----LysGlyLysAsnArgAsn 1485  
QY 1473 AGAAAGGATTTCTAAAAATCACAGATCATTTTCATGAGAGTGCCAAAGCAGAGGACAAAAGA 1532  
Db 1486 ValLysIleGlyLysIleThr-----ValSerGluLysTyr 1497  
QY 1533 AAAGAACAATGTGAAATGAAACATCAAAAGAACAGAAAGGAGATCCCTAAATACATTTCCA 1592  
Db 1498 ArgGlu-----SerValPheArgGlnIleThrAsnAlaAsnGluLeuLysTyrLeuAsp 1515  
QY 1593 CCTCACCCTT-----TCTCCAGATAAGAAATGGCTTGGAACTCTCTTATTGAGGAGATG 1643  
Db 1516 GluPheLeuLeuAsnLysIleAsnAspLeuArgSerGlnLysThrProIleGluSerLeu 1535  
QY 1644 AGGAGAAATGCCAAGGTGTGGGATCCGCTCCCTTGAGACCATTGCTCAATCACACA 1703  
Db 1536 -----PheIleGluAlaThr 1540  
QY 1704 GTGACTATTCGGGTAGATCTTTTGGGAATAGGAGAAGTTTCCTAAA---CCTTTCCCAACA 1760  
Db 1541 GluLysPheArgSerAsnIleLysThrMetTyrSerValProAsnGlyLysIleHisVal 1560  
QY 1761 CATTTTAAAGATTTGTGGGACAAACAGCATGTTAAGATGCCCTTGTTCAGAACAAACTTG 1820  
Db 1561 GlyTyrLysAspLeuMetGluAsnTyrGlnIleValVal-----SerAsnLeu 1576  
QY 1821 TACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACCTCATTCAG 1880  
Db 1577 -----AlaThrGluArgGlyGlnLysAspThrAsnLeuValLeuAsnLeuPheGln 1593  
QY 1881 ACTGCATCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGAT---GCTATTCTGAAG 1937  
Db 1594 ---SerLeuLeuAspGluPheThrArgGlyTyrThrLysAsnAspPheGluProValLys 1612  
QY 1938 TACAATGTGCATATTTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997  
Db 1613 GlnSerLysAlaGlnLysLysLys-----Arg 1621  
QY 1998 GTACTAGAAAGACGAGAGCTCAACAC-----TTGTATCAGTCCATC 2039  
Db 1622 LysGlnGluArgAlaValGlnGluHisAsnGlyHisValPheAlaSerTyrGlnValSer 1641  
QY 2040 TTGCCTGAT-----ATGGTGAATAATGCACCTC 2066  
Db 1642 IleProGlnSerCysGluGlnCysLeuSerTyrIleTyrLeuMetAspLysAlaLeuLeu 1661  
QY 2067 TGTCTGCCAAATATTTGTATCCAGCCCAATACCACTCTGAAACAGAGATGAATCATTC 2126  
Db 1662 Cys-----SerValCys-----Lys 1666  
QY 2127 ATCACAATGTACAGGAACAGATTGCCAGCTCTTTTAGCTAATGCTTTCTTCTGCACGTTT 2186  
Db 1667 MetThrCysHisLysLysCysValHisLysIleGlnSerHisCysSerTyr---ThrTyr 1685  
QY 2187 CCACGACGCAATGCC-----AAGATG 2207  
Db 1686 GlyArgLysGlyGluProGlyAlaGluProGlyHisPheGlyValCysValAspSerLeu 1705  
QY 2208 AAATCAGAGTATTCAGTTATCCAGATATTAATCAATCGTTGTTGAA----- 2258  
Db 1706 ThrSerAspLysAlaSerValPro---IleValLeuGluLysLeuGluHisValGlu 1724  
QY 2259 -----GGACGTTTCATCAAGGAACCA 2279  
Db 1725 MetHisGlyLeuTyrThrGluGlyLeuTyrArgLysSerGlyAlaAlaAsnArgThrArg 1744  
QY 2280 GAGAAGCTTAAACGCTCTTCTGCTACTTTTAGAAGAGTACAGAGAAACCCACTGGG 2339  
Db 1745 Glu-----LeuArgGlnAlaLeuGlnThrAspProAlaAla 1756  
QY 2340 TTGGTGACATTCAAGACAGAGAGTCTTTGAAGATTTTCCA----- 2378

|   |   |     |      |       |   |  |      |
|---|---|-----|------|-------|---|--|------|
| D | b |     | 1757 | Val   | ----  | -----LysLeuGluAsnPheProIleHisAlaIleThrGlyVal | 1770 |
|   |   | ::: |      |       | :::   |  |      |
| Q | y |     | 2379 | ----  | GAGTGG  | -----GAAAGATGTGAATACTCCTGACT                 | 2408 |
|   |   | ::: |      |       |   |  |      |
| D | b |     | 1771 | Leu   | LysGlnTrpLeuArgGluLeuProGluProLeuMetThrPheAlaGlnTyrGlyAsp   | 1790   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2409 | ----  | -----CGACTGCATGTCACT  | 2423   |      |
|   |   | :   |      | :     | :   | :  |      |
| D | b |     | 1791 | Phe   | LeuArgAlaValGluLeuProGluLysGlnGluLeuAlaAlaIleTyrAlaVal      | 1810   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2424 | TAC   | GAAGGTACCATAAGAAGAAACGCCAGGCATGCTACAGGTGGATT                | 2474   |      |
|   |   |     |      |       |   |  |      |
| D | b |     | 1811 | Leu   | GluHisLeuProGluAlaAsnHisAsnSerLeuGluArgLeuIlePheHisLeuVal   | 1830   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2475 | ----  | -----GCAAACCGTTTCGTTGGAGTGTTAACCAGTGCAGGA                   | 2513   |      |
|   |   | :   |      | :     | :   | :  |      |
| D | b |     | 1831 | Lys   | ValAlaLeuLeuGluAspValAsnArg--MetSerProGlyAlaLeuAlaIle       | 1849   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2514 | CTT   | TGTGCAAGAAGAAATCCGCTTTTAATCAACCCCTGAGTTGATTGTTTCACGGCTCTTC  | 2573   |      |
|   |   | :   |      | :     | :   | :  |      |
| D | b |     | 1850 | Phe   | AlaProCysLeuLeuArgCysProAspAsnSerAspProLeuThrSer-----Met    | 1867   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2574 | ACT   | GAGGTGCTGGATCACAATGAATGTTTATCATCACAGGTACTGAGCAGTACAGTGAA    | 2633   |      |
|   |   | :   |      | :     | :   | :  |      |
| D | b |     | 1868 | Lys   | AspValLeuLysIleThrThrCysValGluMetLeuIleLysGluGlnMetArgLys   | 1887   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2634 | TAC   | CACAGGCTATGCCGAAACATACCGCTGGGCCCGAGCCATGAAGACAGGAGCGAAAGG   | 2693   |      |
|   |   |     |      |       |   |  |      |
| D | b |     | 1888 | Tyr   | LysValLysMetGlu-----  | 1893   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2694 | GAC   | GACTGGCAGAGGCGCAGCACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGC   | 2753   |      |
|   |   | :   |      | :     | :   | :  |      |
| D | b |     | 1894 | ----- | -----GluIleSerGlnLeuGluAlaAlaGluSerIleAlaPheArgArg          | 1908   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2754 | ----- | -----TACCTCGACCACTTTGTGCCCGAGAAGATCAGACGGGAGCTTAACAAGGCTTAC | 2807   |      |
|   |   |     |      |       |   |  |      |
| D | b |     | 1909 | Leu   | SerLeuLeuArgGlnAsnAlaProTrpProLeuLysLeuGlyPheSerSerProTyr   | 1928   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2808 | TGT   | GGAATTTCTTCGTCCTGGA-----GTTTCTTCAGAGAACCTGTCT               | 2849   |      |
|   |   |     |      |       |   |  |      |
| D | b |     | 1929 | Glu   | GlyValLeuAsnLysSerProGlnValProArgAspIleGlnGluGluLeuGlu      | 1948   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2850 | GCA   | TGGCTACAGGAAACTGGGCTGTGGTGCCTTTGGGGTGATGCT---AGACTAAA       | 2906   |      |
|   |   | :   |      | :     | :   | :  |      |
| D | b |     | 1949 | Val   | LeuLeuGluGluGlu-----AlaAlaGlyGlyAspGluAspArgGluLys          | 1964   |      |
|   |   | :   |      | :     | :   | :  |      |
| Q | y |     | 2907 | GC    | CTTAATA-----CAGATCCTGGCAGCTGCTGTAGTCAGCGAGACGTGGTTTTAT      | 2957   |      |
|   |   | : : |      | :     | :   | :  |      |
| D | b |     | 1965 | Glu   | IleLeuIleGluArgIleGlnSerIleLysGluGluLysGluAspIleThrTyr      | 1983   |      |
|   |   | :   |      | :     | :   | :  |      |





; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 11  
; LENGTH: 31  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-11

Query Match 100.0%; Score 167; DB 4; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.6e-18;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIITGTEQYSEYTGTAETYS 31  
Db 1 LFTVLDHNECLIITGTEQYSEYTGTAETYS 31

RESULT 3  
US-09-511-507-11  
; Sequence 11, Application US/09511507  
; Patent No. 6395543

; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 11  
; LENGTH: 31  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-11

Query Match 100.0%; Score 167; DB 4; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.6e-18;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIITGTEQYSEYTGTAETYS 31  
Db 1 LFTVLDHNECLIITGTEQYSEYTGTAETYS 31

RESULT 4  
US-09-302-812-6  
; Sequence 6, Application US/09302812B  
; Patent No. 6333148

; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT

; ORGANISM: Mus musculus  
; FEATURE:  
US-09-302-812-6

Query Match 100.0%; Score 167; DB 4; Length 968;  
Best Local Similarity 100.0%; Pred. No. 9.5e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIITGTEQYSEYTGTAETYS 31  
Db 762 LFTVLDHNECLIITGTEQYSEYTGTAETYS 792

RESULT 5  
US-09-511-477-6  
; Sequence 6, Application US/09511477  
; Patent No. 6337202

; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-511-477-6

Query Match 100.0%; Score 167; DB 4; Length 968;  
Best Local Similarity 100.0%; Pred. No. 9.5e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIITGTEQYSEYTGTAETYS 31  
Db 762 LFTVLDHNECLIITGTEQYSEYTGTAETYS 792

RESULT 6  
US-09-511-507-6  
; Sequence 6, Application US/09511507  
; Patent No. 6395543

; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-511-507-6

Query Match 100.0%; Score 167; DB 4; Length 968;



; Sequence 2, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-2

Query Match 100.0%; Score 167; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 9.6e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LFTVLDHNECLIIITGTEQYSEYTGTAETR 31  
Db 771 LFTVLDHNECLIIITGTEQYSEYTGTAETR 801

RESULT 12  
US-09-511-507-2  
; Sequence 2, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-2

Query Match 100.0%; Score 167; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 9.6e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LFTVLDHNECLIIITGTEQYSEYTGTAETR 31  
Db 771 LFTVLDHNECLIIITGTEQYSEYTGTAETR 801

RESULT 13  
US-09-302-812-8  
; Sequence 8, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-302-812-8

Query Match 49.1%; Score 82; DB 4; Length 768;  
Best Local Similarity 53.3%; Pred. No. 0.00051;  
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 LFTVLDHNECLIIITGTEQYSEYTGTAETY 30  
Db 400 LFTVLDHNECLIIITGTEQYSEYTGTAETY 429

RESULT 14  
US-09-511-477-8  
; Sequence 8, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-477-8

Query Match 49.1%; Score 82; DB 4; Length 768;  
Best Local Similarity 53.3%; Pred. No. 0.00051;  
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 LFTVLDHNECLIIITGTEQYSEYTGTAETY 30  
Db 400 LFTVLDHNECLIIITGTEQYSEYTGTAETY 429

RESULT 15  
US-09-511-507-8  
; Sequence 8, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH



Search completed: May 26, 2004, 18:49:30  
Job time : 0.993349 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 3.55169 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-11  
Perfect score: 167  
Sequence: 1 LFTVELDHNECLLIITGTEQYSEYTGVAETYS 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A Geneseq\_29Jan04:\*\*
- 1: Geneseqp1980s:\*\*
  - 2: Geneseqp1990s:\*\*
  - 3: Geneseqp2000s:\*\*
  - 4: Geneseqp2001s:\*\*
  - 5: Geneseqp2002s:\*\*
  - 6: Geneseqp2003as:\*\*
  - 7: Geneseqp2003bs:\*\*
  - 8: Geneseqp2004s:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description        |
|------------|-------|---------------|--------|-------|--------------------|
| 1          | 167   | 100.0         | 31     | 5     | Aae25634 Bovine po |
| 2          | 167   | 100.0         | 31     | 5     | Aae25651 Bovine po |
| 3          | 167   | 100.0         | 31     | 5     | Aau76025 Bovine po |
| 4          | 167   | 100.0         | 31     | 5     | Aau76016 Bovine po |
| 5          | 167   | 100.0         | 31     | 6     | Abg72283 Oligopept |
| 6          | 167   | 100.0         | 968    | 5     | Aae25631 Murine po |
| 7          | 167   | 100.0         | 968    | 5     | Aau76022 Mouse pol |
| 8          | 167   | 100.0         | 968    | 5     | Aau76013 Mouse pol |
| 9          | 167   | 100.0         | 968    | 6     | Abg72280 Mmurine p |
| 10         | 167   | 100.0         | 976    | 5     | Aae25630 Human pol |
| 11         | 167   | 100.0         | 976    | 5     | Aau76021 Human pol |
| 12         | 167   | 100.0         | 976    | 5     | Aau76012 Human pol |
| 13         | 167   | 100.0         | 976    | 6     | Abg72279 Human pol |
| 14         | 167   | 100.0         | 977    | 5     | Aae25629 Bovine po |
| 15         | 167   | 100.0         | 977    | 5     | Aau76020 Bovine po |
| 16         | 167   | 100.0         | 977    | 5     | Aau75799 Bovine po |
| 17         | 167   | 100.0         | 977    | 6     | Abg72278 Bovine po |
| 18         | 97    | 58.1          | 819    | 4     | Abg20721 Novel hum |
| 19         | 82    | 49.1          | 768    | 4     | Abb59491 Drosophil |
| 20         | 82    | 49.1          | 768    | 5     | Aae25632 Fruit fly |
| 21         | 82    | 49.1          | 768    | 5     | Aau76023 Fruit fly |
| 22         | 82    | 49.1          | 768    | 5     | Aau76014 Fruit fly |
| 23         | 82    | 49.1          | 768    | 6     | Abg72281 Fruit fly |
| 24         | 54.5  | 32.6          | 428    | 6     | Abu24640 Protein e |
| 25         | 53    | 31.7          | 726    | 5     | Aae25633 Poly aden |

|    |      |      |      |   |          |                    |
|----|------|------|------|---|----------|--------------------|
| 26 | 53   | 31.7 | 726  | 5 | AAU76024 | Aau76024 Worm poly |
| 27 | 53   | 31.7 | 726  | 5 | AAU76015 | Aau76015 Worm poly |
| 28 | 53   | 31.7 | 726  | 6 | ABG72282 | Abg72282 C. elegan |
| 29 | 52   | 31.1 | 464  | 4 | AAB86542 | Aab86542 E. rhacon |
| 30 | 50   | 29.9 | 172  | 1 | AAP93508 | Aap93508 Sequence  |
| 31 | 50   | 29.9 | 259  | 3 | AAG16762 | Aag16762 Arabidops |
| 32 | 49.5 | 29.6 | 123  | 7 | ADC95672 | Adc95672 E. faeciu |
| 33 | 49.5 | 29.6 | 307  | 6 | ABU44051 | Abu44051 Protein e |
| 34 | 49.5 | 29.6 | 434  | 7 | ADC94448 | Adc94448 E. faeciu |
| 35 | 49   | 29.3 | 447  | 6 | ABU44183 | Abu44183 Protein e |
| 36 | 49   | 29.3 | 453  | 2 | AAR86293 | Aar86293 Protamino |
| 37 | 49   | 29.3 | 3138 | 4 | ABB61958 | Abb61958 Drosophil |
| 38 | 48.5 | 29.0 | 101  | 5 | ABB10079 | Abb10079 Human Hly |
| 39 | 48.5 | 29.0 | 122  | 7 | ADC96224 | Adc96224 E. faeciu |
| 40 | 48   | 28.7 | 372  | 2 | AAW22175 | Aaw22175 S.thermop |
| 41 | 48   | 28.7 | 384  | 2 | AAW14076 | Aaw14076 S.thermop |
| 42 | 48   | 28.7 | 993  | 5 | AAE17835 | Aae17835 Herpes si |
| 43 | 48   | 28.7 | 1037 | 5 | AAE17837 | Aae17837 Herpes si |
| 44 | 48   | 28.7 | 1057 | 2 | AAW72067 | Aaw72067 HSV-2 str |
| 45 | 48   | 28.7 | 1113 | 5 | AAE17836 | Aae17836 Herpes si |

ALIGNMENTS

RESULT 1  
AAE25634  
ID AAE25634 standard; peptide; 31 AA.  
XX  
AC AAE25634;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #1.  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
PN US63955543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson BL, Ame J, Lin W;  
XX  
XX WPI; 2002-535641/57.  
XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Claim 10; Col 81-82; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing complementary

CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is bPARG oligopeptide  
XX  
SQ Sequence 31 AA;  
  
Query Match 100.0%; Score 167; DB 5; Length 31;  
Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LFTEVLDRNECLIIITGTEQYSEYTGTAETYSR 31  
DB 1 LFTEVLDRNECLIIITGTEQYSEYTGTAETYSR 31  
  
RESULT 2  
AAE25651  
ID AAE25651 standard; peptide; 31 AA.  
XX  
AC AAE25651;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #7.  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Example 3; Col 27; 77pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX The present sequence is bPARG oligopeptide

SQ Sequence 31 AA;  
  
Query Match 100.0%; Score 167; DB 5; Length 31;  
Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LFTEVLDRNECLIIITGTEQYSEYTGTAETYSR 31  
DB 1 LFTEVLDRNECLIIITGTEQYSEYTGTAETYSR 31  
  
RESULT 3  
AAU76025  
ID AAU76025 standard; peptide; 31 AA.  
XX  
AC AAU76025;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 68.  
XX  
KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease; oligopeptide 68.  
XX  
OS Bos taurus.  
XX  
PN US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-163240/21.  
XX  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
PS Claim 4; Col 25; 81pp; English.  
XX  
XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents bovine PARG  
CC oligonucleotide 68. This peptide is one of several PARG oligopeptides  
CC (AAU76025-AAU76028) of the invention  
XX  
SQ Sequence 31 AA;  
  
Query Match 100.0%; Score 167; DB 5; Length 31;  
Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIIITGTEQYSEYTGTAETIR 31  
DE |||||  
XX 1 LFTVLDHNECLIIITGTEQYSEYTGTAETIR 31  
RESULT 4  
AAU76016  
ID AAU76016 standard; peptide; 31 AA.  
XX  
AC AAU76016;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 68.  
XX  
KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity; oligopeptide 68.  
XX  
OS Bos taurus.  
XX US6333148-B1.  
PN  
XX 25-DEC-2001.  
PD  
XX 30-APR-1999; 99US-00302812.  
PF  
XX 01-MAY-1998; 98US-0083768P.  
PR  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-153820/20.  
XX  
XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
PS Claim 4; Col 25; 80pp; English.  
XX  
XX The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage,  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents bovine PARG oligopeptide 68. This peptide is one of several  
CC PARG oligopeptides (AAU76016-AAU76019) of the invention  
XX  
SQ Sequence 31 AA;  
Query Match 100.0%; Score 167; DB 5; Length 31;  
Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVLDHNECLIIITGTEQYSEYTGTAETIR 31  
DE |||||  
1 LFTVLDHNECLIIITGTEQYSEYTGTAETIR 31  
RESULT 5  
ABG72283  
ID ABG72283 standard; peptide; 31 AA.  
XX  
AC ABG72283;  
XX

DT 13-MAR-2003 (first entry)  
XX  
DE Oligopeptide #1 derived from bovine PARG enzyme.  
XX  
KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective.  
XX  
OS Bos taurus.  
XX US2002132328-A1.  
PN  
XX 19-SEP-2002.  
PD  
XX 09-OCT-2001; 2001US-00973451.  
PF  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2003-155895/15.  
DR  
XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
XX Example 2; Page 14; 86pp; English.  
PS  
XX The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. ABG72283-  
CC ABG72286 represent oligopeptides derived from bovine PARG enzyme. The  
CC oligopeptides are used to construct degenerate PCR primers for the  
CC isolation of cDNA encoding bovine PARG  
XX  
SQ Sequence 31 AA;  
Query Match 100.0%; Score 167; DB 6; Length 31;  
Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVLDHNECLIIITGTEQYSEYTGTAETIR 31  
DE |||||  
1 LFTVLDHNECLIIITGTEQYSEYTGTAETIR 31  
RESULT 6  
AAE25631  
ID AAE25631 standard; protein; 968 AA.



XX AAE25631;  
AC 04-NOV-2002 (first entry)  
XX Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
DE  
XX Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytotstatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX Mus musculus.  
OS US63955543-B1.  
PN 28-MAY-2002.  
XX 23-FEB-2000; 2000US-00511507.  
PF 01-MAY-1998; 98US-0083768P.  
XX 30-APR-1999; 99US-00302812.  
PR (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-535641/57.  
DR N-PSDB; AAD42083.  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX Claim 3; Col 63-68; 77pp; English.  
XX The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is murine PARG  
XX Sequence 968 AA;  
SQ Query Match 100.0%; Score 167; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTEVLHDHNECLIIITGTEQYSEYTGGAETVR 31  
Db 762 LFTEVLHDHNECLIIITGTEQYSEYTGGAETVR 792  
RESULT 7  
AAU76022  
ID AAU76022 standard; protein; 968 AA.  
XX AAU76022;  
AC 08-MAY-2002 (first entry)  
XX Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX Mus musculus.  
OS US6337202-B1.  
PN 08-JAN-2002.  
XX 23-FEB-2000; 2000US-00511477.  
PF 01-MAY-1998; 98US-0083768P.  
XX 30-APR-1999; 99US-00302812.  
PR (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-163240/21.  
DR N-PSDB; ABK14933.  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX Claim 2; Col 63-70; 81pp; English.  
XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the mouse PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX Sequence 968 AA;  
SQ Query Match 100.0%; Score 167; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTEVLHDHNECLIIITGTEQYSEYTGGAETVR 31  
Db 762 LFTEVLHDHNECLIIITGTEQYSEYTGGAETVR 792  
RESULT 8  
AAU76013  
ID AAU76013 standard; protein; 968 AA.  
XX AAU76013;  
AC 08-MAY-2002 (first entry)  
XX Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;



PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-005111507.  
XX  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-535641/57.  
DR N-PSDB; AAD42082.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
XX Claim 3; Col 55-60; 77pp; English.  
PS  
XX The invention relates to an isolated nucleic acid molecule which encodes  
XX a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is human PARG  
XX  
XX Sequence 976 AA;  
SQ  
Query Match 100.0%; Score 167; DB 5; Length 976;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTEVLHDNECLIIITGTEQYSEYTGTAETTYR 31  
DB 770 LFTEVLHDNECLIIITGTEQYSEYTGTAETTYR 800  
RESULT 11  
AAU76021  
ID AAU76021 standard; protein; 976 AA.  
XX  
XX AAU76021;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
XX Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
XX Homo sapiens.  
XX  
XX US6337202-B1.  
XX  
XX 08-JAN-2002.  
XX  
XX 23-FEB-2000; 2000US-005111477.  
XX  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-163240/21.  
DR N-PSDB; ABK14932.  
XX  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
XX Claim 2; Col 55-60; 81pp; English.  
PS  
XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the human PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX  
XX Sequence 976 AA;  
SQ  
Query Match 100.0%; Score 167; DB 5; Length 976;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTEVLHDNECLIIITGTEQYSEYTGTAETTYR 31  
DB 770 LFTEVLHDNECLIIITGTEQYSEYTGTAETTYR 800  
RESULT 12  
AAU76012  
ID AAU76012 standard; protein; 976 AA.  
XX  
XX AAU76012;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX  
XX Homo sapiens.  
XX  
XX US6333148-B1.  
XX  
XX 25-DEC-2001.  
XX  
XX 30-APR-1999; 99US-00302812.  
XX  
XX 01-MAY-1998; 98US-0083768P.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-153820/20.  
DR N-PSDB; ABK14494.

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
PS Claim 3; Col 55-60; 80pp; English.  
XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage,  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the human PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX  
SQ Sequence 976 AA;  
Query Match 100.0%; Score 167; DB 5; Length 976;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTEVLHDHNECLIIITGTEQYSEYTGAEYR 31  
Db 770 LFTEVLHDHNECLIIITGTEQYSEYTGAEYR 800  
RESULT 13  
ID ABG72279 standard; protein; 976 AA.  
XX  
AC ABG72279;  
XX  
DT 13-MAR-2003 (first entry)  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX  
KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; neurotropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective; enzyme.  
XX  
OS Homo sapiens.  
XX  
PN US2002132328-A1.  
XX  
PD 19-SEP-2002.  
XX  
PF 09-OCT-2001; 2001US-00973451.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2003-155895/15.  
DR N-PSDB; ABX14478.  
XX  
PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)

PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
PS Claim 28; Fig 16; 86pp; English.  
XX  
CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC diseases (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents human PARG enzyme  
XX  
SQ Sequence 976 AA;  
Query Match 100.0%; Score 167; DB 6; Length 976;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTEVLHDHNECLIIITGTEQYSEYTGAEYR 31  
Db 770 LFTEVLHDHNECLIIITGTEQYSEYTGAEYR 800  
RESULT 14  
ID AAE25629 standard; protein; 977 AA.  
XX  
AC AAE25629;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytosstatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-535641/57.  
DR N-PSDB; AAD42081.  
XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.



XX XX

Search completed: May 26, 2004, 18:40:08  
Job time : 4.55169 secs



US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

Query Match 100.0%; Score 167; DB 9; Length 977;  
Best Local Similarity 100.0%; Pred. No. 1.8e-16;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LFTEVLDPNECLIIITGTEQYSEYTGAEYR 31  
Db 771 LFTEVLDPNECLIIITGTEQYSEYTGAEYR 801

RESULT 5  
US-09-973-451-8  
; Sequence 8, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-973-451-8

Query Match 49.1%; Score 82; DB 9; Length 768;  
Best Local Similarity 53.3%; Pred. No. 0.0013;  
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
  
QY 1 LFTEVLDPNECLIIITGTEQYSEYTGAEY 30  
Db 400 LFTEVLDPNECLIIITGTEQYSEYTGAGSF 429

US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

Query Match 100.0%; Score 167; DB 9; Length 968;  
Best Local Similarity 100.0%; Pred. No. 1.7e-16;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LFTEVLDPNECLIIITGTEQYSEYTGAEYR 31  
Db 762 LFTEVLDPNECLIIITGTEQYSEYTGAEYR 792

RESULT 3  
US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4

Query Match 100.0%; Score 167; DB 9; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.8e-16;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LFTEVLDPNECLIIITGTEQYSEYTGAEYR 31  
Db 770 LFTEVLDPNECLIIITGTEQYSEYTGAEYR 800

```
RESULT 6
US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-262-E12_FLI.pep
US-10-425-114-60000

Query Match      43.1%; Score 72; DB 12; Length 546;
Best Local Similarity 38.7%; Pred. No. 0.029;
Matches 12; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY      1 LFTEVLHDNECLITGTEQYSEYTGAEYR 31
|| : : : : : : : : : : : : : : : : : :
Db      330 LFLSCMEDNEAIEIFGAERFSQYMGYGSSFR 360

RESULT 7
US-10-424-599-233915
; Sequence 233915, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 233915
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_5324C.1.pep
US-10-424-599-233915

Query Match      33.5%; Score 56; DB 12; Length 300;
Best Local Similarity 41.7%; Pred. No. 3.9;
Matches 10; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY      8 HNECLITGTEQYSEYTGAEYR 31
: : : : : : : : : : : : : : : : : :
Db      216 YNERIIVGVVERFSGYTDHASSFR 239

RESULT 8
US-10-425-114-65345
; Sequence 65345, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
```

```
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65345
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4759-008-F1_FLI.pep
US-10-425-114-65345

Query Match      32.9%; Score 55; DB 12; Length 524;
Best Local Similarity 47.6%; Pred. No. 11;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      7 DHNECLITGTEQYSEYTGGA 27
|| : : : : : : : : : : : : : : : : : :
Db      446 DHDMCVPTGTGTEAWTASLGGA 466

RESULT 9
US-10-425-114-65514
; Sequence 65514, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65514
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4767-001-H9_FLI.pep
US-10-425-114-65514

Query Match      32.9%; Score 55; DB 12; Length 529;
Best Local Similarity 47.6%; Pred. No. 11;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      7 DHNECLITGTEQYSEYTGGA 27
|| : : : : : : : : : : : : : : : : : :
Db      451 DHDMCVPTGTGTEAWTASLGGA 471

RESULT 10
US-10-282-122A-52564
; Sequence 52564, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
```



```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52564
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52564

Query Match      32.6%; Score 54.5; DB 12; Length 428;
Best Local Similarity 38.9%; Pred. No. 10;
Matches 14; Conservative 4; Mismatches 5; Indels 13; Gaps 2;

QY    1 LFTEV-----LDHNECLIITGTEQYSEYTGVAETYS 31
       : ||: ||: ||: ||: ||: ||: ||: ||: ||: |
Db    172 IITEIFGRALDYNEI-----YNYLGYAEKLR 199

RESULT 11
US-09-973-451-10
; Sequence 10, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
```

; Sequence 45983, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 45983  
; LENGTH: 224  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700894080\_FLI.pep  
US-10-425-114-45983

Query Match 29.3%; Score 49; DB 12; Length 224;  
Best Local Similarity 40.0%; Pred. No. 32;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 7 DHNECLIIITGTEQYSEYTG 26  
||: |: ||: |: ||  
Db 146 DHDMCVPTGTSEAWTRSLGY 165

RESULT 14  
US-10-425-114-52216  
; Sequence 52216, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 52216  
; LENGTH: 329  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700788604\_FLI.pep  
US-10-425-114-52216

Query Match 29.3%; Score 49; DB 12; Length 329;  
Best Local Similarity 40.0%; Pred. No. 51;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 7 DHNECLIIITGTEQYSEYTG 26  
||: |: ||: |: ||  
Db 251 DHDMCVPTGTSEAWTRSLGY 270

RESULT 15  
US-10-389-566-1764  
; Sequence 1764, Application US/10389566  
; Publication No. US20040025202A1  
; GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology, LLC  
; APPLICANT: Laurie, Cathy C

; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants  
; FILE REFERENCE: 38-77(52900)D  
; CURRENT APPLICATION NUMBER: US/10/389,566  
; CURRENT FILING DATE: 2003-03-31  
; PRIOR APPLICATION NUMBER: US 60/365,301  
; PRIOR FILING DATE: 2002-03-15  
; PRIOR APPLICATION NUMBER: US 60/391,786  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: US 60/392,018  
; PRIOR FILING DATE: 2002-06-26  
; NUMBER OF SEQ ID NOS: 2459  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1764  
; LENGTH: 427  
; TYPE: PRT  
; ORGANISM: Trichodesmium erythraeum  
US-10-389-566-1764

Query Match 29.3%; Score 49; DB 16; Length 427;  
Best Local Similarity 36.0%; Pred. No. 70;  
Matches 9; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 5 VLDHNECLIIITGTEQYSEYTG 29  
||: |: ||: |: ||  
Db 254 ILELECALSRGARIYAEMVGYGMT 278

Search completed: May 26, 2004, 19:19:00  
Job time : 3.58646 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 0.496675 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-11  
Perfect score: 167  
Sequence: 1 LFTVLDHNECLITGTQYSEYTGAEYR 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues 141681  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID        | Description        |
|------------|-------|-------------|--------|--------------|--------------------|
| 1          | 52    | 31.1        | 510    | 1 CBP1_ORYSA | P37890 oryza sativ |
| 2          | 50.5  | 30.2        | 541    | 1 YB36_METJA | Q58536 methanococc |
| 3          | 50    | 29.9        | 364    | 1 VPAP_HSV7J | P52440 human herpe |
| 4          | 50    | 29.9        | 499    | 1 CBP1_HORVU | P07519 hordeum vul |
| 5          | 48.5  | 29.0        | 558    | 1 LEU1_CLOAB | Q97mc5 clostridium |
| 6          | 48    | 28.7        | 250    | 1 YM25_ARCFU | O28058 archaeoglob |
| 7          | 48    | 28.7        | 474    | 1 VP61_NPVOP | O10270 orgyia pseu |
| 8          | 48    | 28.7        | 521    | 1 NU2C_SYNPF | P29801 synechococc |
| 9          | 47.5  | 28.4        | 844    | 1 MCEL_VACCC | P20979 vaccinia vi |
| 10         | 47.5  | 28.4        | 844    | 1 MCEL_VACCV | P04298 vaccinia vi |
| 11         | 47.5  | 28.4        | 2186   | 1 YL52_CAEEL | P34431 caenorhabdi |
| 12         | 47    | 28.1        | 520    | 1 NU2C_ANASP | Q8ymq0 anabaena sp |
| 13         | 47    | 28.1        | 593    | 1 RECG_PASMU | Q9cmb4 pasteurella |
| 14         | 46.5  | 27.8        | 164    | 1 PHEA_PORPU | P51368 porphyra pu |
| 15         | 46.5  | 27.8        | 164    | 1 PHEA_PORTE | O49843 porphyra te |
| 16         | 46.5  | 27.8        | 164    | 1 PHEA_PORVE | O20206 porphyra ye |
| 17         | 46.5  | 27.8        | 164    | 1 PHEA_RHOVL | Q02036 rhodella vi |
| 18         | 46.5  | 27.8        | 401    | 1 ASSY_STAAM | Q99vc7 staphylococ |
| 19         | 46.5  | 27.8        | 401    | 1 ASSY_STAAM | Q8nxf2 staphylococ |
| 20         | 46.5  | 27.8        | 796    | 1 COPP_SCHPO | O42937 schizosacch |
| 21         | 46.5  | 27.8        | 2863   | 1 LRBA_HUMAN | P50851 homo sapien |
| 22         | 46    | 27.5        | 76     | 1 CSBA_BACSU | P37953 bacillus su |
| 23         | 46    | 27.5        | 518    | 1 SP5B_BACSU | Q00758 bacillus su |
| 24         | 46    | 27.5        | 527    | 1 SYK_CHLPN  | Q09265 chlamydia p |
| 25         | 45.5  | 27.2        | 1020   | 1 YRD3_CAEEL | Q09573 caenorhabdi |
| 26         | 45    | 26.9        | 168    | 1 DYP_BACSU  | P11045 bacillus su |
| 27         | 45    | 26.9        | 254    | 1 CTRL_HALRU | P35003 haliotis ru |
| 28         | 45    | 26.9        | 354    | 1 GB12_CHICK | P50147 gallus gall |
| 29         | 45    | 26.9        | 415    | 1 GLYA_LACLA | Q9chw7 lactococcus |
| 30         | 45    | 26.9        | 450    | 1 SQRD_MOUSE | Q9rl12 mus musculu |
| 31         | 45    | 26.9        | 511    | 1 MA12_PENCI | P31723 penicillium |
| 32         | 45    | 26.9        | 556    | 1 TCPA_SCHPO | O94501 schizosacch |
| 33         | 45    | 26.9        | 708    | 1 CAO4_CANMA | P05335 candida mal |

|    |      |      |      |              |                    |
|----|------|------|------|--------------|--------------------|
| 34 | 45   | 26.9 | 723  | 1 CAO4_CANTR | P11356 candida tro |
| 35 | 45   | 26.9 | 797  | 1 AF32_HUMAN | Q9y4w6 homo sapien |
| 36 | 45   | 26.9 | 1040 | 1 RIK1_SCHPO | Q10426 schizosacch |
| 37 | 45   | 26.9 | 1390 | 1 INSR_AEDAE | Q93105 aedes aegyp |
| 38 | 44.5 | 26.6 | 598  | 1 FLIF_PSEAE | Q51463 pseudomonas |
| 39 | 44.5 | 26.6 | 742  | 1 ZW10_ARATH | O48626 arabidopsis |
| 40 | 44.5 | 26.6 | 1000 | 1 COPP_CAEEL | Q20168 caenorhabdi |
| 41 | 44   | 26.3 | 97   | 1 CH10_BUCTC | Q9f4e8 buchnera ap |
| 42 | 44   | 26.3 | 137  | 1 COPI_DICDI | P54706 dictyosteli |
| 43 | 44   | 26.3 | 161  | 1 UREE_PROMI | P17090 proteus mir |
| 44 | 44   | 26.3 | 164  | 1 PHA1_SYNPW | Q08086 synechococc |
| 45 | 44   | 26.3 | 164  | 1 PHA1_SYNPY | Q02179 synechococc |

ALIGNMENTS

|  |  |           |      |         |  |
|--|--|-----------|------|---------|--|
| RESULT 1   |  |           |      |         |  |
| CBP1_ORYSA   |  |           |      |         |  |
| ID   | CBP1_ORYSA   | STANDARD; | PRT; | 510 AA. |  |
| AC   | P37890;  |           |      |         |  |
| DT   | 01-OCT-1994 (Rel. 30, Created)   |           |      |         |  |
| DT   | 01-OCT-1994 (Rel. 30, Last sequence update)  |           |      |         |  |
| DT   | 10-OCT-2003 (Rel. 42, Last annotation update)  |           |      |         |  |
| DE   | Serine carboxypeptidase I precursor (EC 3.4.16.5) (Carboxypeptidase C).              |           |      |         |  |
| GN   | CBP1.  |           |      |         |  |
| OS   | Oryza sativa (Rice).   |           |      |         |  |
| OC   | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;                   |           |      |         |  |
| OC   | Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;                           |           |      |         |  |
| OC   | Ehrhartoideae; Oryzeae; Oryza.   |           |      |         |  |
| OX   | NCBI_TaxID=4530;   |           |      |         |  |
| RN   | [1]  |           |      |         |  |
| RP   | SEQUENCE FROM N.A.   |           |      |         |  |
| RC   | STRAIN=cv. Yukihikari;   |           |      |         |  |
| RX   | MEDLINE=94213891; PubMed=8161571;  |           |      |         |  |
| RA   | Washio K., Ishikawa K.;  |           |      |         |  |
| RT   | "Cloning and sequencing of the gene for type I carboxypeptidase in rice."            |           |      |         |  |
| RL   | Biochim. Biophys. Acta 1199:311-314(1994).   |           |      |         |  |
| CC   | -I- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a broad specificity. |           |      |         |  |
| CC   | -I- PTM: Three disulfide bonds are present (Potential).                              |           |      |         |  |
| CC   | -I- SIMILARITY: Belongs to peptidase family S10.                                     |           |      |         |  |
| -----  |  |           |      |         |  |
| This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ). |  |           |      |         |  |
| -----  |  |           |      |         |  |
| EMBL;  | D17586; BAA04510.1; --   |           |      |         |  |
| PIR;   | S43516; S43516.  |           |      |         |  |
| HSSP;  | P08819; 1WHI.  |           |      |         |  |
| MEROPS;  | S10.004; --  |           |      |         |  |
| Gramene;   | P37890; --   |           |      |         |  |
| InterPro;  | IPR001563; Peptidase_S10.  |           |      |         |  |
| InterPro;  | IPR000379; Ser esters.   |           |      |         |  |
| Pfam;  | PF00450; serine carboxpept; 1.   |           |      |         |  |
| PRINTS;  | PR00724; CRBOXYPTASEC.   |           |      |         |  |
| ProDom;  | PD001189; Serine carbpept; 2.  |           |      |         |  |
| PROSITE;   | PS00131; CARBOXYPEPT_SER_SER; 1.   |           |      |         |  |
| PROSITE;   | PS00560; CARBOXYPEPT_SER_HIS; 1.   |           |      |         |  |
| KW   | Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.                          |           |      |         |  |
| FT   | SIGNAL 1 25  |           |      |         |  |
| FT   | PROPEP 26 36   |           |      |         |  |
| FT   | CHAIN 37 510   |           |      |         |  |
| FT   | ACT_SITE 194 194   |           |      |         |  |
| FT   | ACT_SITE 434 434   |           |      |         |  |
| FT   | ACT_SITE 487 487   |           |      |         |  |
| FT   | CARBOHYD 154 154   |           |      |         |  |
| N-LINKED (GLCNAC. . .) (POTENTIAL).  |  |           |      |         |  |

```

FT CARBOHYD      268      268      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD      418      418      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE          508      510      MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE      510 AA; 55709 MW; 28896247FA1371CF CRC64;

Query Match      31.1%; Score 52; DB 1; Length 510;
Best Local Similarity 45.0%; Pred. No. 8.2;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 7 DHNECLIIITGTEQYSEYTG Y 26
DB 432 DHDMCVPTGTETWTRSLGY 451

RESULT 2
YB36 METJA
ID YB36 METJA STANDARD; PRT; 541 AA.
AC Q58536;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1136.
GN MJ1136.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Huret M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii.";
SC Science 273:1058-1073(1996).
CC -! SIMILARITY: STRONG, TO C.ELEGANS ZK863.3 AND S.CEREVISIAE LPG22.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67556; AAB99138.1; -.
DR PIR; G64441; G64441.
DR TIGR; MJ1136; -.
DR InterPro; IPR005910; ELP3.
DR InterPro; IPR006638; Elp3.
DR InterPro; IPR000182; GCN5acetyl_trans.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF00583; Acetyltransf; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMS; TIGR01211; ELP3; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 541 AA; 62443 MW; 31794B06F11E0BFD CRC64;

Query Match      30.2%; Score 50.5; DB 1; Length 541;
Best Local Similarity 30.2%; Pred. No. 14;
Matches 13; Conservative 6; Mismatches 9; Indels 15; Gaps 2

QY 1 LFTEVLHDHNE-----CLIIITGTEQY-----SEYTGAYE 28
:
```





```
CC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- SIMILARITY: Belongs to the FAH family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000952; AAB89031.1; -.
CC PIR; A69528; A69528.
CC TIGR; AF2225; -.
CC InterPro; IPR002529; FAA hydrolase.
CC Pfam; PF01557; FAA hydrolase; 1.
CC Hypothetical protein; Complete proteome.
KW SEQUENCE 250 AA; 27370 MW; AD0E196CA19C6EC0 CRC64;
SQ
Query Match 28.7%; Score 48; DB 1; Length 250;
Best Local Similarity 34.8%; Pred. No. 14;
Matches 8; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
QY 3 TEVLHDNECLITGTQYSEYTG 25
Db 76 TAVIGHDDCILLPQISQVDYEG 98
RESULT 7
VP61 NPVOP
ID VP61 NPVOP STANDARD; PRT; 474 AA.
AC O10270;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 61 kDa protein homolog.
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
RN [2]
RP REVISIONS.
RA Rohrmann G.F.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Probably plays an important role in the persistence and
CC survival of the virus. May be a structural component in the
```

```
CC envelope of the polyhedron or the envelope of the polyhedra-
CC derived virus (PDV) (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75930; AAC59001.1; -.
CC InterPro; IPR003124; WH2.
CC Pfam; PF02205; WH2; 2.
CC SMART; SM00246; WH2; 2.
CC DOMAIN 239 330 PRO-RICH.
CC DOMAIN 267 273 POLY-PRO.
CC DOMAIN 282 289 POLY-PRO.
CC DOMAIN 298 302 POLY-PRO.
CC DOMAIN 311 319 POLY-PRO.
CC SEQUENCE 474 AA; 52009 MW; B8A5EA4C3BC0D236 CRC64;
SQ
Query Match 28.7%; Score 48; DB 1; Length 474;
Best Local Similarity 31.2%; Pred. No. 28;
Matches 10; Conservative 10; Mismatches 10; Indels 2; Gaps 1;
QY 1 LFTEVLHDNECLITGT--EQYSEYTGAEY 30
Db 116 LRVDVNDAEVNVLSGDFEYYSKYISYQQT 147
RESULT 8
NU2C SYN7
ID NU2C SYN7 STANDARD; PRT; 521 AA.
AC P29801;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NAD(P)H-quinone oxidoreductase chain 2 (EC 1.6.5.-) (NAD(P)H
DE dehydrogenase I, chain 2) (NDH-1, chain 2).
GN NDHB.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94143466; PubMed=8310046;
RA Marco E., Chad N., Schwarz R., Liemen-Hurwitz J., Gabay C., Kaplan A.;
RT "High CO2 concentration alleviates the block in photosynthetic
RT electron transport in an ndhB-inactivated mutant of Synechococcus sp.
RT PCC 7942.";
RL Plant Physiol. 101:1047-1053(1993).
CC -!- FUNCTION: NDH-1 shuttles electrons from NAD(P)H, via FMN and iron-
CC sulfur (Fe-S) centers, to quinones in the respiratory chain. The
CC immediate electron acceptor for the enzyme in this species is
CC believed to be plastoquinone. Couples the redox reaction to proton
CC translocation (for every two electrons transferred, four hydrogen
CC ions are translocated across the cytoplasmic membrane), and thus
CC conserves the redox energy in a proton gradient (By similarity).
CC -!- CATALYTIC ACTIVITY: NAD(P)H + plastoquinone = NAD(P) (+) +
CC plastoquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the complex I subunit 2 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X65027; CAA46161.1; -.
```











Search completed: May 26, 2004, 18:41:17  
Job time : 2.49667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 2.474 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-11  
Perfect score: 167  
Sequence: 1 LFTEVLHDHNECLITGTEQYSEYTGAEYR 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID     | Description        |
|------------|-------|---------------|--------|-----------|--------------------|
| 1          | 167   | 100.0         | 920    | 11 Q8CB72 | Q8cb72 mus musculu |
| 2          | 167   | 100.0         | 961    | 11 Q80YQ6 | Q80yq6 mus musculu |
| 3          | 167   | 100.0         | 968    | 11 Q88622 | Q88622 mus musculu |
| 4          | 167   | 100.0         | 972    | 11 Q9QYM2 | Q9qym2 rattus norv |
| 5          | 167   | 100.0         | 976    | 4 Q9Y4W7  | Q9y4w7 homo sapien |
| 6          | 167   | 100.0         | 976    | 4 Q86W56  | Q86w56 homo sapien |
| 7          | 167   | 100.0         | 976    | 4 Q7Z742  | Q7z742 homo sapien |
| 8          | 167   | 100.0         | 977    | 6 Q02776  | Q02776 bos taurus  |
| 9          | 82    | 49.1          | 723    | 5 Q960N8  | Q960n8 drosophila  |
| 10         | 82    | 49.1          | 768    | 5 Q46043  | Q46043 drosophila  |
| 11         | 79    | 47.3          | 548    | 10 Q9SKB3 | Q9skb3 arabidopsis |
| 12         | 69    | 41.3          | 522    | 10 Q8VYA1 | Q8vya1 arabidopsis |
| 13         | 57    | 34.1          | 368    | 5 Q86GI4  | Q86gi4 toxoplasma  |
| 14         | 55    | 32.9          | 485    | 5 Q9N5L4  | Q9n5l4 caenorhabdi |
| 15         | 53.5  | 32.0          | 437    | 17 Q8TR71 | Q8tr71 methanosarc |
| 16         | 53.5  | 32.0          | 941    | 16 Q8DFP9 | Q8dfp9 vibrio vuln |

|    |      |      |      |           |                    |
|----|------|------|------|-----------|--------------------|
| 17 | 53   | 31.7 | 764  | 5 Q19637  | Q19637 caenorhabdi |
| 18 | 53   | 31.7 | 781  | 5 Q867X0  | Q867x0 caenorhabdi |
| 19 | 52   | 31.1 | 464  | 2 Q9AI65  | Q9ai65 erwinia rha |
| 20 | 51   | 30.5 | 132  | 16 Q32092 | Q32092 bacillus su |
| 21 | 51   | 30.5 | 238  | 11 Q9D410 | Q9d410 mus musculu |
| 22 | 50.5 | 30.2 | 332  | 5 Q86GZ6  | Q86gz6 rhipicephal |
| 23 | 50   | 29.9 | 166  | 16 Q8ZC80 | Q8zce0 yersinia pe |
| 24 | 50   | 29.9 | 364  | 10 Q9SKB4 | Q9skb4 arabidopsis |
| 25 | 50   | 29.9 | 456  | 10 Q9SV78 | Q9sv78 arabidopsis |
| 26 | 50   | 29.9 | 497  | 10 Q8L7B2 | Q8l7b2 arabidopsis |
| 27 | 50   | 29.9 | 1759 | 5 Q95PL4  | Q95pl4 trypanosoma |
| 28 | 49   | 29.3 | 447  | 16 Q8DTC3 | Q8dte3 streptococc |
| 29 | 49   | 29.3 | 504  | 10 Q9LSV8 | Q9lsv8 arabidopsis |
| 30 | 49   | 29.3 | 618  | 2 Q9F5E5  | Q9f5e5 agrobacteri |
| 31 | 49   | 29.3 | 2523 | 5 Q9VTP0  | Q9vtp0 drosophila  |
| 32 | 48   | 28.7 | 364  | 2 Q8GCY0  | Q8gcy0 bartonella  |
| 33 | 48   | 28.7 | 372  | 2 Q56044  | Q56044 streptococc |
| 34 | 48   | 28.7 | 384  | 2 Q8KIC9  | Q8kic9 streptococc |
| 35 | 48   | 28.7 | 392  | 2 Q9LCI0  | Q9lci0 streptomyce |
| 36 | 48   | 28.7 | 413  | 16 Q8RGX4 | Q8rgx4 fusobacteri |
| 37 | 48   | 28.7 | 427  | 2 Q9FAR5  | Q9far5 streptomyce |
| 38 | 48   | 28.7 | 515  | 16 Q8DMR6 | Q8dmr6 synechococc |
| 39 | 48   | 28.7 | 659  | 16 Q816J3 | Q816j3 bacillus ce |
| 40 | 48   | 28.7 | 1114 | 12 P89460 | P89460 herpes simp |
| 41 | 48   | 28.7 | 1925 | 5 Q8I2D1  | Q8i2d1 plasmodium  |
| 42 | 47.5 | 28.4 | 276  | 16 Q8F4H5 | Q8f4h5 leptospira  |
| 43 | 47.5 | 28.4 | 441  | 2 Q84FV7  | Q84fv7 methylobact |
| 44 | 47.5 | 28.4 | 843  | 12 Q8JLC5 | Q8jlc5 ectromelia  |
| 45 | 47.5 | 28.4 | 844  | 12 Q57209 | Q57209 vaccinia vi |

ALIGNMENTS

RESULT 1

Q8CB72 ID Q8CB72 PRELIMINARY; PRT; 920 AA.  
AC Q8CB72;  
DT 01-MAR-2003 (TremBLrel. 23, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)  
DE Poly.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
DR EMBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 100.0%; Score 167; DB 11; Length 920;  
Best Local Similarity 100.0%; Pred. NO. 2.5e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTEVLHDHNECLITGTEQYSEYTGAEYR 31

Db 763 LFTEVLHDHNECLITGTEQYSEYTGAEYR 793

RESULT 2

Q80YQ6





Q86W56  
ID Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TReMBLrel. 24, Created)  
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
  
Query Match 100.0%; Score 167; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 2.7e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LfTEVLDHNECLIITGTEQYSEYTGyaETyr 31  
Db 770 LfTEVLDHNECLIITGTEQYSEYTGyaETyr 800  
  
RESULT 7  
Q7Z742  
ID Q7Z742 PRELIMINARY; PRT; 976 AA.  
AC Q7Z742;  
DT 01-OCT-2003 (TReMBLrel. 25, Created)  
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;  
RA Klausner R.D.; Collins F.S.; Wagner L.; Shenmen C.M.; Schuler G.D.;  
RA Altschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K.;  
RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Hsieh F.;  
RA Diatchenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;  
RA Stapleton M.; Soares M.B.; Bonaldo M.F.; Casavant T.L.; Scheetz T.E.;  
RA Brownstein M.J.; Usdin T.B.; Toshiyuki S.; Carninci P.; Prange C.;  
RA Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mullahy S.J.;  
RA Bosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;  
RA Richards S.; Worley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Hulyk S.W.;  
RA Villalon D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.;  
RA Fahey J.; Helton E.; Kettman M.; Madan A.; Rodrigues S.; Sanchez A.;  
RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;  
RA Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;  
RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield Y.S.;  
RA Krzywinski M.I.; Skalska U.; Smailus D.E.; Schnerch A.; Schein J.E.;  
RA Jones S.J.; Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC052966; AAH52966.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;  
  
Query Match 100.0%; Score 167; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 2.7e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LfTEVLDHNECLIITGTEQYSEYTGyaETyr 31  
Db 770 LfTEVLDHNECLIITGTEQYSEYTGyaETyr 800  
  
RESULT 8  
O02776  
ID O02776 PRELIMINARY; PRT; 977 AA.  
AC O02776;  
DT 01-JUL-1997 (TReMBLrel. 04, Created)  
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN BPARG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97277328; PubMed=9115250;  
RA Lin W.; Ame J.C.; Aboul-Ela N.; Jacobson E.L.; Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-  
RT ribose) glycohydrolase."  
RL J. Biol. Chem. 272:11895-11901(1997).  
DR EMBL; U78975; AAB53370.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;  
  
Query Match 100.0%; Score 167; DB 6; Length 977;  
Best Local Similarity 100.0%; Pred. No. 2.7e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LfTEVLDHNECLIITGTEQYSEYTGyaETyr 31  
Db 771 LfTEVLDHNECLIITGTEQYSEYTGyaETyr 801  
  
RESULT 9  
Q960N8  
ID Q960N8 PRELIMINARY; PRT; 723 AA.  
AC Q960N8;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE LD42380p.  
GN PARG OR EG:114E2.1 OR CG2864.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M.; Brokstein P.; Hong L.; Agbayani A.; Carlson J.;  
RA Champe M.; Chavez C.; Dorsett V.; Farfan D.; Frise E.; George R.;  
RA Gonzalez M.; Guarin H.; Li P.; Liao G.; Miranda A.; Mungall C.J.;  
RA Nunoo J.; Pacleb J.; Paragas V.; Park S.; Phouanavong S.; Wan K.;  
RA Yu C.; Lewis S.E.; Rubin G.M.; Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY051955; AAK93379.1; -.  
DR FlyBase; FBgn0023216; PARG.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;  
  
Query Match 49.1%; Score 82; DB 5; Length 723;  
Best Local Similarity 53.3%; Pred. No. 0.0038;  
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
  
QY 1 LFTEVLHDHNECLITGTGEQYSEYTGAEY 30  
Db 355 LFTECLRPFEALVLMGAERYSNYTGAGSF 384  
  
RESULT 10  
O46043 ID O46043 PRELIMINARY; PRT; 768 AA.  
AC O46043;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE PARG protein.  
GN PARG OR EG:114E2.1 OR CG2864.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.

RA Ame J.-C., Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding Drosophila  
RT poly(ADP-ribose) glycohydrolase."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 46-768 FROM N.A.  
RA Murphy L., Harris D., Barrell B.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 46-768 FROM N.A.  
RA Benos P.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003428; AAF45886.1; -.  
DR EMBL; AF079556; AAC28734.1; -.  
DR EMBL; Z98254; CAB10913.1; -.  
DR FlyBase; FBgn0023216; PARG.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;  
  
Query Match 49.1%; Score 82; DB 5; Length 768;  
Best Local Similarity 53.3%; Pred. No. 0.0041;  
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
  
QY 1 LFTEVLHDHNECLITGTGEQYSEYTGAEY 30  
Db 400 LFTECLRPFEALVLMGAERYSNYTGAGSF 429  
  
RESULT 11  
Q9SKB3 ID Q9SKB3 PRELIMINARY; PRT; 548 AA.  
AC Q9SKB3; Q94ET7;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative poly(ADP-ribose) glycohydrolase.  
GN AT2G31870 OR TEJ.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E., M.-I.,  
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,  
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,  
RA Fraser C.M., Venter J.C.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Panda S., Poirier G.G., Kay S.A.;  
RT "TEJ defines a role for poly-ADP-ribosylation in establishing period  
RT length of the Arabidopsis circadian oscillator."  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006533; AAD32285.2; -.  
DR EMBL; AF394690; AAK72256.1; -.  
DR PIR; B84726; B84726.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 548 AA; 62169 MW; F1A79FDA157C3329 CRC64;  
  
Query Match 47.3%; Score 79; DB 10; Length 548;  
Best Local Similarity 48.4%; Pred. No. 0.0077;  
Matches 15; Conservative 6; Mismatches 10; Indels 0; Gaps 0;  
  
QY 1 LFTEVLHDHNECLITGTGEQYSEYTGAEY 31



OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;  
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2214;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
RX MEDLINE=21929760; PubMed=11932238;  
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,  
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,  
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,  
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
RA Metcalf W.W., Birren B.;  
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
RT and physiological diversity.";  
RL Genome Res. 12:532-542(2002).  
DR EMBL; AE010800; AAM04729.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 437 AA; 48229 MW; BB4A089F86593F5E CRC64;

Query Match 32.0%; Score 53.5; DB 17; Length 437;  
Best Local Similarity 46.2%; Pred. No. 29;  
Matches 12; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

QY 1 LFTVEVDHNECLIIITGTEQYSEYTG 26  
:| |||| : : |||| : : ||  
Db 172 VFEEVLD-ADLIISTGTVEFHYAGY 196

Search completed: May 26, 2004, 18:46:08  
Job time : 4.474 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 3.32255 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-12  
Perfect score: 152  
Sequence: 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 150   | 98.7        | 29     | 5     | AAE25635    |
| 2          | 150   | 98.7        | 29     | 5     | AAU76026    |
| 3          | 150   | 98.7        | 29     | 5     | AAU76017    |
| 4          | 150   | 98.7        | 29     | 6     | ABG72284    |
| 5          | 149   | 98.0        | 976    | 5     | AAE25630    |
| 6          | 149   | 98.0        | 976    | 5     | AAU76021    |
| 7          | 149   | 98.0        | 976    | 5     | AAU76012    |
| 8          | 149   | 98.0        | 976    | 6     | ABG72279    |
| 9          | 149   | 98.0        | 977    | 5     | AAE25629    |
| 10         | 149   | 98.0        | 977    | 5     | AAU76020    |
| 11         | 149   | 98.0        | 977    | 5     | AAU75799    |
| 12         | 149   | 98.0        | 977    | 6     | ABG72278    |
| 13         | 144   | 94.7        | 33     | 5     | AAE25652    |
| 14         | 144   | 94.7        | 968    | 5     | AAE25631    |
| 15         | 144   | 94.7        | 968    | 5     | AAU76022    |
| 16         | 144   | 94.7        | 968    | 5     | AAU76013    |
| 17         | 144   | 94.7        | 968    | 6     | ABG72280    |
| 18         | 69    | 45.4        | 768    | 4     | ABB59491    |
| 19         | 69    | 45.4        | 768    | 5     | AAE25632    |
| 20         | 69    | 45.4        | 768    | 5     | AAU76023    |
| 21         | 69    | 45.4        | 768    | 5     | AAU76014    |
| 22         | 69    | 45.4        | 768    | 6     | ABG72281    |
| 23         | 60.5  | 39.8        | 543    | 2     | AAR42456    |
| 24         | 60.5  | 39.8        | 543    | 2     | AAR99465    |
| 25         | 60.5  | 39.8        | 543    | 2     | AAW37053    |

|    |      |      |     |   |          |
|----|------|------|-----|---|----------|
| 26 | 60.5 | 39.8 | 543 | 2 | AAW89403 |
| 27 | 60.5 | 39.8 | 543 | 3 | AAB10470 |
| 28 | 55   | 36.2 | 578 | 4 | ABB71910 |
| 29 | 53   | 34.9 | 75  | 5 | ABU67233 |
| 30 | 53   | 34.9 | 75  | 6 | ABP54014 |
| 31 | 53   | 34.9 | 403 | 3 | AAU44642 |
| 32 | 53   | 34.9 | 403 | 3 | AAU90673 |
| 33 | 53   | 34.9 | 403 | 3 | AAU90638 |
| 34 | 53   | 34.9 | 403 | 4 | AAB99185 |
| 35 | 53   | 34.9 | 403 | 5 | ABG95158 |
| 36 | 53   | 34.9 | 403 | 5 | ABG95171 |
| 37 | 53   | 34.9 | 403 | 6 | ABP81682 |
| 38 | 53   | 34.9 | 403 | 6 | ADA99010 |
| 39 | 53   | 34.9 | 403 | 7 | ADC22633 |
| 40 | 53   | 34.9 | 403 | 7 | ADC22743 |
| 41 | 53   | 34.9 | 403 | 7 | ADE31653 |
| 42 | 53   | 34.9 | 426 | 4 | AAE03628 |
| 43 | 53   | 34.9 | 426 | 5 | AAE15631 |
| 44 | 53   | 34.9 | 445 | 4 | AAU68566 |
| 45 | 53   | 34.9 | 445 | 4 | AAU68523 |

ALIGNMENTS

RESULT 1  
AAE25635  
ID AAE25635 standard; peptide; 29 AA.  
XX

AC AAE25635;

XX  
DT 04-NOV-2002 (first entry)

XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #2.  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytotatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
XX  
OS Bos taurus.

XX  
FH Key Location/Qualifiers  
FT Misc-difference 23  
FT /label= Unknown

XX  
PN US6395543-B1.

XX  
PD 28-MAY-2002.

XX  
PF 23-FEB-2000; 2000US-00511507.

XX  
PR 01-MAY-1998; 98US-0083768P.

XX  
PR 30-APR-1999; 99US-00302812.

XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.

XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;

XX  
XX WPI; 2002-535641/57.

XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX  
PS Claim 10; Col 81-82; 77pp; English.

XX  
CC The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP ribose polymer. The

CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is bPARG oligopeptide

XX  
SQ Sequence 29 AA;

Query Match 98.7%; Score 150; DB 5; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
|||||  
Db 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

RESULT 2  
AAU76026  
ID AAU76026 standard; peptide; 29 AA.

XX  
AC AAU76026;

XX 08-MAY-2002 (first entry)

DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 63.

XX  
KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease; oligopeptide 63.

OS Bos taurus.

XX Key Location/Qualifiers

FT Misc-difference 23 /label= Unknown  
FT /note= "Xaa is not further defined in the specification"

XX US6337202-B1.

XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-00511477.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-163240/21.

XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 4; Col 25; 81pp; English.

XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,

CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents bovine PARG  
CC oligonucleotide 63. This peptide is one of several PARG oligopeptides  
XX (AAU76025-AAU76028) of the invention

SQ Sequence 29 AA;

Query Match 98.7%; Score 150; DB 5; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
|||||  
Db 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

RESULT 3  
AAU76017  
ID AAU76017 standard; peptide; 29 AA.

XX  
AC AAU76017;

XX 08-MAY-2002 (first entry)

DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 63.

XX  
KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity; oligopeptide 63.

OS Bos taurus.

XX Key Location/Qualifiers

FT Misc-difference 23 /label= Unknown  
FT /note= "Xaa is not further defined in the specification"

XX US6333148-B1.

XX 25-DEC-2001.

XX 30-APR-1999; 99US-00302812.

XX 01-MAY-1998; 98US-0083768P.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-153820/20.

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.

XX Claim 4; Col 25; 80pp; English.

XX The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known

CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents bovine PARG oligopeptide 63. This peptide is one of several  
CC PARG oligopeptides (AAU76016-AAU76019) of the invention  
XX  
SQ Sequence 29 AA;  
  
Query Match 98.7%; Score 150; DB 5; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
DB 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
  
RESULT 4  
ABG72284  
ID ABG72284 standard; peptide; 29 AA.  
XX  
AC ABG72284;  
XX  
DT 13-MAR-2003 (first entry)  
DE Oligopeptide #2 derived from bovine PARG enzyme.  
XX  
KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective.  
XX  
OS Bos taurus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 23 /label= Unknown  
FT  
XX  
PN US2002132328-A1.  
XX  
PD 19-SEP-2002.  
XX  
PF 09-OCT-2001; 2001US-00973451.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2003-155895/15.  
XX  
PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
PS Example 2; Page 14; 86pp; English.  
XX  
CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic

CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. ABG72283-  
CC ABG72286 represent oligopeptides derived from bovine PARG enzyme. The  
CC oligopeptides are used to construct degenerate PCR primers for the  
CC isolation of cDNA encoding bovine PARG  
XX  
SQ Sequence 29 AA;  
  
Query Match 98.7%; Score 150; DB 6; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
DB 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
  
RESULT 5  
AAE25630  
ID AAE25630 standard; protein; 976 AA.  
XX  
AC AAE25630;  
XX  
DT 04-NOV-2002 (first entry)  
XX Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
DE  
XX Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisenese therapy.  
XX  
OS Homo sapiens.  
XX  
PN US63955543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-535641/57.  
DR N-PSDB; AAD42082.  
XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Claim 3; Col 55-60; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress

CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is human PARG  
XX  
SQ Sequence 976 AA;  
  
Query Match 98.0%; Score 149; DB 5; Length 976;  
Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876  
  
RESULT 6  
AAU76021  
ID AAU76021 standard; protein; 976 AA.  
XX  
AC AAU76021;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
OS Homo sapiens.  
XX  
PN US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-163240/21.  
DR N-PSDB; ABK14932.  
XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
PS Claim 2; Col 55-60; 81pp; English.  
XX  
CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the human PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX

SQ Sequence 976 AA;  
  
Query Match 98.0%; Score 149; DB 5; Length 976;  
Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876  
  
RESULT 7  
AAU76012  
ID AAU76012 standard; protein; 976 AA.  
XX  
AC AAU76012;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX  
OS Homo sapiens.  
XX  
PN US6333148-B1.  
XX  
PD 25-DEC-2001.  
XX  
PF 30-APR-1999; 99US-00302812.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-153820/20.  
DR N-PSDB; ABK14494.  
XX  
PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
PS Claim 3; Col 55-60; 80pp; English.  
XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the human PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX  
SQ Sequence 976 AA;  
  
Query Match 98.0%; Score 149; DB 5; Length 976;  
Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876



RESULT 8  
ABG72279  
ID ABG72279 standard; protein; 976 AA.  
XX  
AC ABG72279;  
XX  
DT 13-MAR-2003 (first entry)  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX  
KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective; enzyme.  
XX  
OS Homo sapiens.  
XX  
PN US2002132328-A1.  
XX  
PD 19-SEP-2002.  
XX  
PF 09-OCT-2001; 2001US-00973451.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
WPI; 2003-155895/15.  
DR N-PSDB; ABX14478.  
XX  
PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
PS Claim 28; Fig 16; 86pp; English.  
XX  
CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents human PARG enzyme  
XX  
SQ Sequence 976 AA;  
Query Match 98.0%; Score 149; DB 6; Length 976;  
Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGFLRPGVSSNLSAVATGNXGCGAFG 29

Db 848 AYCGFLRPGVSSNLSAVATGNWCGGAFG 876  
|||||  
RESULT 9  
AAE25629  
ID AAE25629 standard; protein; 977 AA.  
XX  
AC AAE25629;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
WPI; 2002-535641/57.  
DR N-PSDB; AAD42081.  
XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Claim 3; Col 47-45; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX  
SQ Sequence 977 AA;  
Query Match 98.0%; Score 149; DB 5; Length 977;  
Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGFLRPGVSSNLSAVATGNXGCGAFG 29  
|||||  
Db 849 AYCGFLRPGVSSNLSAVATGNWCGGAFG 877  
|||||  
RESULT 10  
AAU76020

|  |   |
|--|---|
| ID   | AAU76020 standard; protein; 977 AA.   |
| XX AC  | AAU76020;   |
| XX DT  | 08-MAY-2002 (first entry)   |
| XX DE  | Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.   |
| XX KW  | Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose; adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma; hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke; neurodegenerative disease; neurological disorder; Alzheimer's disease; Huntington's disease; Parkinson's disease.  |
| XX OS  | Bos taurus.   |
| XX FH  | Key Location/Qualifiers   |
| FT Region  | 601..617  |
| FT FT  | /note= "Represents PARG oligopeptide #1"  |
| FT Region  | 761..770  |
| FT FT  | /note= "Represents PARG oligopeptide #2"  |
| FT Region  | 771..801  |
| FT FT  | /note= "Represents PARG oligopeptide #3"  |
| FT Region  | 849..880  |
| FT FT  | /note= "Represents PARG oligopeptide #4"  |
| XX PN  | US6337202-B1.   |
| XX PD  | 08-JAN-2002.  |
| XX PF  | 23-FEB-2000; 2000US-00511477.   |
| XX PR  | 01-MAY-1998; 98US-0083768P.   |
| XX PR  | 30-APR-1999; 99US-00302812.   |
| XX PA  | (KENT ) UNIV KENTUCKY RES FOUND.  |
| XX PI  | Jacobson MK, Jacobson EL, Ame J, Lin W;   |
| XX DR  | WPI; 2002-163240/21.  |
| XX DR  | N-PSDB; ABK14931.   |
| XX PT  | Novel isolated and purified poly (ADP-ribose) glycohydrolase protein which catalyses release of ADP-ribose from ADP ribose polymer, useful for treating neoplastic and neurological disorders, heart attack and stroke.   |
| XX PS  | Claim 2; Col 47-52; 81pp; English.  |
| XX CC  | The present invention relates to a new poly(ADP-ribose) glycohydrolase (PARG) protein which catalyses release of ADP-ribose from an ADP (adenosine diphosphate)-ribose polymer. The PARG molecule of the invention is useful for generating antibodies and can be inhibited or activated for diagnosing and treating neoplastic disorders such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, hyperplasia and hypertrophy, reperfusion following ischaemia, heart attack, stroke, neurodegenerative diseases, neurological disorders including Alzheimer's, Huntington's and Parkinson's diseases, and related conditions. PARG levels may be enhanced to suppress DNA repair and increase the cell's susceptibility to chemotherapy drugs. Antagonists of PARG are administered to treat or prevent neoplastic disorders. The present amino acid sequence represents the bovine PARG protein of the invention. This protein is one of several PARG proteins (AAU76020-AAU76024) of the invention |
| XX SQ  | Sequence 977 AA;  |
| Query Match 98.0%; Score 149; DB 5; Length 977;            |   |
| Best Local Similarity 96.6%; Pred. No. 1.8e-12;            |   |
| Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0 |   |
| OY   | 1 AYCGFLRPGVSSENLSAVATGNXGCCGAFG 29<br>   |

Db 849 AYCGLRPGVSSSENLSAVATGNWCGAFG 877  
 RESULT 11  
 AAU75799  
 ID AAU75799 standard; protein; 977 AA.  
 XX  
 AC AAU75799;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX  
 OS Bos taurus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 601..617  
 FT /note= "Represents PARG oligopeptide #1"  
 FT Region 761..770  
 FT /note= "Represents PARG oligopeptide #2"  
 FT Region 771..801  
 FT /note= "Represents PARG oligopeptide #3"  
 FT Region 849..880  
 FT /note= "Represents PARG oligopeptide #4"  
 XX  
 PN US6333148-B1.  
 XX  
 PD 25-DEC-2001.  
 XX  
 PF 30-APR-1999; 99US-00302812.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 XX  
 PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2002-153820/20.  
 DR N-PSDB; ABK14493.  
 XX  
 PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX  
 PS Claim 3; Col 45-52; 80pp; English.  
 XX  
 CC The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage,  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the bovine PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX  
 SQ Sequence 977 AA;  
 Query Match 98.0%; Score 149; DB 5; Length 977;  
 Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AYCGLRPGVSSSENLSAVATGNWCGAFG 29

Db 849 AYCGLRPGVSSSENLSAVATGNWCGGAFG 877  
|||||  
RESULT 12  
ABG72278  
ID ABG72278 standard; protein; 977 AA.  
XX  
AC ABG72278;  
XX  
DT 13-MAR-2003 (first entry)  
XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX  
KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; neotropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective; enzyme.  
XX  
OS Bos taurus.  
XX  
PN US2002132328-A1.  
XX  
PD 19-SEP-2002.  
XX  
PF 09-OCT-2001; 2001US-00973451.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2003-155895/15.  
DR N-PSDB; ABX14477.  
XX  
PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
PS Claim 28; Fig 16; 86pp; English.  
XX  
CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents bovine PARG enzyme  
XX  
SQ Sequence 977 AA;  
Query Match 98.0%; Score 149; DB 6; Length 977;  
Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
|||||  
Db 849 AYCGLRPGVSSSENLSAVATGNWCGGAFG 877  
|||||  
RESULT 13  
AAE25652  
ID AAE25652 standard; peptide; 33 AA.  
XX  
AC AAE25652;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #8.  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-535641/57.  
XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Example 3; Col 27; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX  
SQ Sequence 33 AA;  
Query Match 94.7%; Score 144; DB 5; Length 33;  
Best Local Similarity 93.1%; Pred. No. 2.6e-13;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
|||||  
Db 1 AYCGLRPGVPSNLSAVATGNWCGGAFG 29  
|||||

RESULT 14

AAE25631  
ID AAE25631 standard; protein; 968 AA.  
XX  
AC AAE25631;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
XX  
KW Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cyostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Mus musculus.  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-535641/57.  
DR N-PSDB; AAD42083.  
XX  
XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Claim 3; Col 63-68; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is murine PARG  
XX  
SQ Sequence 968 AA;  
  
Query Match 94.7%; Score 144; DB 5; Length 968;  
Best Local Similarity 93.1%; Pred. No. 9.2e-12;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
Db 840 AYCGLRPGVPSNLSAVATGNWGGCGAFG 868  
  
RESULT 15  
AAU76022  
ID AAU76022 standard; protein; 968 AA.  
XX  
AC AAU76022;  
XX  
DT 08-MAY-2002 (first entry)

XX  
DE Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Mouse; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
OS Mus musculus.  
XX  
PN US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-163240/21.  
DR N-PSDB; ABK14933.  
XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
PS Claim 2; Col 63-70; 81pp; English.  
XX  
CC The present invention relates to a new poly (ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the mouse PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX  
SQ Sequence 968 AA;  
  
Query Match 94.7%; Score 144; DB 5; Length 968;  
Best Local Similarity 93.1%; Pred. No. 9.2e-12;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
Db 840 AYCGLRPGVPSNLSAVATGNWGGCGAFG 868  
  
Search completed: May 26, 2004, 18:40:09  
Job time : 4.32255 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 0.929262 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-12  
Perfect score: 152  
Sequence: 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Query Match % | Length | DB ID              | Description       |
|------------|-------|-------|---------------|--------|--------------------|-------------------|
| 1          | 150   | 98.7  | 29            | 4      | US-09-302-812-12   | Sequence 12, Appl |
| 2          | 150   | 98.7  | 29            | 4      | US-09-511-477-12   | Sequence 12, Appl |
| 3          | 150   | 98.7  | 29            | 4      | US-09-511-507-12   | Sequence 12, Appl |
| 4          | 149   | 98.0  | 976           | 4      | US-09-302-812-4    | Sequence 4, Appli |
| 5          | 149   | 98.0  | 976           | 4      | US-09-511-477-4    | Sequence 4, Appli |
| 6          | 149   | 98.0  | 976           | 4      | US-09-511-507-4    | Sequence 4, Appli |
| 7          | 149   | 98.0  | 977           | 4      | US-09-302-812-2    | Sequence 2, Appli |
| 8          | 149   | 98.0  | 977           | 4      | US-09-511-477-2    | Sequence 2, Appli |
| 9          | 149   | 98.0  | 977           | 4      | US-09-511-507-2    | Sequence 2, Appli |
| 10         | 144   | 94.7  | 968           | 4      | US-09-302-812-6    | Sequence 6, Appli |
| 11         | 144   | 94.7  | 968           | 4      | US-09-511-477-6    | Sequence 6, Appli |
| 12         | 144   | 94.7  | 968           | 4      | US-09-511-507-6    | Sequence 6, Appli |
| 13         | 69    | 45.4  | 768           | 4      | US-09-302-812-8    | Sequence 8, Appli |
| 14         | 69    | 45.4  | 768           | 4      | US-09-511-477-8    | Sequence 8, Appli |
| 15         | 69    | 45.4  | 768           | 4      | US-09-511-507-8    | Sequence 8, Appli |
| 16         | 60.5  | 39.8  | 543           | 1      | US-08-375-709-17   | Sequence 17, Appl |
| 17         | 60.5  | 39.8  | 543           | 1      | US-08-752-929-17   | Sequence 17, Appl |
| 18         | 60.5  | 39.8  | 543           | 3      | US-09-090-793-10   | Sequence 10, Appl |
| 19         | 60.5  | 39.8  | 543           | 4      | US-09-231-899-10   | Sequence 10, Appl |
| 20         | 53    | 34.9  | 403           | 4      | US-09-170-496D-114 | Sequence 114, App |
| 21         | 53    | 34.9  | 403           | 4      | US-09-170-496D-224 | Sequence 224, App |
| 22         | 53    | 34.9  | 403           | 4      | US-09-743-742B-4   | Sequence 4, Appli |
| 23         | 53    | 34.9  | 403           | 4      | US-09-743-742B-10  | Sequence 10, Appl |
| 24         | 50.5  | 33.2  | 411           | 4      | US-09-540-236-3549 | Sequence 3549, Ap |
| 25         | 49.5  | 32.6  | 726           | 4      | US-09-302-812-10   | Sequence 10, Appl |
| 26         | 49.5  | 32.6  | 726           | 4      | US-09-511-477-10   | Sequence 10, Appl |
| 27         | 49.5  | 32.6  | 726           | 4      | US-09-511-507-10   | Sequence 10, Appl |

|    |      |      |      |   |                      |                   |
|----|------|------|------|---|----------------------|-------------------|
| 28 | 49   | 32.2 | 1196 | 4 | US-09-252-991A-18261 | Sequence 18261, A |
| 29 | 48.5 | 31.9 | 887  | 4 | US-09-077-940A-2     | Sequence 2, Appli |
| 30 | 47.5 | 31.2 | 298  | 3 | US-09-025-691-1      | Sequence 1, Appli |
| 31 | 46.5 | 30.6 | 134  | 4 | US-09-621-976-4066   | Sequence 4066, Ap |
| 32 | 46.5 | 30.6 | 307  | 3 | US-09-049-672A-9     | Sequence 9, Appli |
| 33 | 46   | 30.3 | 359  | 2 | US-08-713-636-2      | Sequence 2, Appli |
| 34 | 46   | 30.3 | 681  | 1 | US-08-083-590A-18    | Sequence 18, Appl |
| 35 | 46   | 30.3 | 681  | 2 | US-08-346-128-37     | Sequence 37, Appl |
| 36 | 46   | 30.3 | 681  | 3 | US-08-532-384-18     | Sequence 18, Appl |
| 37 | 46   | 30.3 | 1068 | 1 | US-08-537-210A-2     | Sequence 2, Appli |
| 38 | 46   | 30.3 | 1068 | 3 | US-09-113-825-2      | Sequence 2, Appli |
| 39 | 46   | 30.3 | 1078 | 1 | US-08-264-534-32     | Sequence 32, Appl |
| 40 | 46   | 30.3 | 1078 | 1 | US-08-083-590A-11    | Sequence 11, Appl |
| 41 | 46   | 30.3 | 1078 | 1 | US-08-465-500-32     | Sequence 32, Appl |
| 42 | 46   | 30.3 | 1078 | 2 | US-08-346-128-32     | Sequence 32, Appl |
| 43 | 46   | 30.3 | 1078 | 3 | US-08-532-384-11     | Sequence 11, Appl |
| 44 | 46   | 30.3 | 1078 | 3 | US-08-893-828-32     | Sequence 32, Appl |
| 45 | 46   | 30.3 | 2556 | 1 | US-08-185-432-17     | Sequence 17, Appl |

ALIGNMENTS

RESULT 1  
US-09-302-812-12  
; Sequence 12, Application US/09302812B  
; Patent No. 6333148.  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 12  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-12  
Query Match 98.7%; Score 150; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 1.2e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
Db 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
RESULT 2  
US-09-511-477-12  
; Sequence 12, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812

```
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 12
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-477-12

Query Match          98.7%; Score 150; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

RESULT 3
US-09-511-507-12
; Sequence 12, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 12
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-507-12

Query Match          98.7%; Score 150; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

RESULT 4
US-09-302-812-4
; Sequence 4, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
; FEATURE:
US-09-302-812-4

Query Match          98.0%; Score 149; DB 4; Length 976;
Best Local Similarity 96.6%; Pred. No. 7e-13;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db      848 AYCGLRPGVSSSENLSAVATGNWCGGAFG 876

RESULT 5
US-09-511-477-4
; Sequence 4, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-511-477-4

Query Match          98.0%; Score 149; DB 4; Length 976;
Best Local Similarity 96.6%; Pred. No. 7e-13;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db      848 AYCGLRPGVSSSENLSAVATGNWCGGAFG 876

RESULT 6
US-09-511-507-4
; Sequence 4, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-511-507-4

Query Match          98.0%; Score 149; DB 4; Length 976;
Best Local Similarity 96.6%; Pred. No. 7e-13;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db      848 AYCGLRPGVSSSENLSAVATGNWCGGAFG 876

RESULT 6
US-09-511-507-4
; Sequence 4, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-511-507-4

Query Match          98.0%; Score 149; DB 4; Length 976;
```

```
Best Local Similarity 96.6%; Pred. No. 7e-13;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSLSAVATGNXGCGAFG 29
Db 848 AYCGLRPGVSSSLSAVATGNWCGGAFG 876

RESULT 7
US-09-302-812-2
; Sequence 2, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-302-812-2

Query Match 98.0%; Score 149; DB 4; Length 977;
Best Local Similarity 96.6%; Pred. No. 7e-13;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSLSAVATGNXGCGAFG 29
Db 849 AYCGLRPGVSSSLSAVATGNWCGGAFG 877

RESULT 8
US-09-511-477-2
; Sequence 2, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-477-2

Query Match 98.0%; Score 149; DB 4; Length 977;
Best Local Similarity 96.6%; Pred. No. 7e-13;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSLSAVATGNXGCGAFG 29
Db 840 AYCGLRPGVSSSLSAVATGNWCGGAFG 868

RESULT 11
US-09-511-477-6
```

```
Db 849 AYCGLRPGVSSSLSAVATGNWCGGAFG 877

RESULT 9
US-09-511-507-2
; Sequence 2, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-507-2

Query Match 98.0%; Score 149; DB 4; Length 977;
Best Local Similarity 96.6%; Pred. No. 7e-13;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSLSAVATGNXGCGAFG 29
Db 849 AYCGLRPGVSSSLSAVATGNWCGGAFG 877

RESULT 10
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-302-812-6

Query Match 94.7%; Score 144; DB 4; Length 968;
Best Local Similarity 93.1%; Pred. No. 3.5e-12;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSLSAVATGNXGCGAFG 29
Db 840 AYCGLRPGVSSSLSAVATGNWCGGAFG 868

RESULT 11
US-09-511-477-6
```

```
; Sequence 6, Application US/095111477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
;
US-09-511-477-6

Query Match      94.7%; Score 144; DB 4; Length 968;
Best Local Similarity 93.1%; Pred. No. 3.5e-12;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
      |||||
Db      840 AYCGLRPGVPSSENLSAVATGNWGCAGFG 868

RESULT 12
US-09-511-507-6
; Sequence 6, Application US/095111507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
;
US-09-511-507-6

Query Match      94.7%; Score 144; DB 4; Length 968;
Best Local Similarity 93.1%; Pred. No. 3.5e-12;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
      |||||
Db      840 AYCGLRPGVPSSENLSAVATGNWGCAGFG 868

RESULT 13
US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
```

```
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
;
US-09-302-812-8

Query Match      45.4%; Score 69; DB 4; Length 768;
Best Local Similarity 48.6%; Pred. No. 0.12;
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY      1 AYCGLR-----PGVSSSENLSAVATGNXGCGAFG 29
      |||||
Db      478 AYIGFVHMVTPPPG-----VATGNWGCAGFG 504

RESULT 14
US-09-511-477-8
; Sequence 8, Application US/095111477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
;
US-09-511-477-8

Query Match      45.4%; Score 69; DB 4; Length 768;
Best Local Similarity 48.6%; Pred. No. 0.12;
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY      1 AYCGLR-----PGVSSSENLSAVATGNXGCGAFG 29
      |||||
Db      478 AYIGFVHMVTPPPG-----VATGNWGCAGFG 504

RESULT 15
US-09-511-507-8
; Sequence 8, Application US/095111507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
```



```

; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

```

|                       |              |                 |               |             |
|-----------------------|--------------|-----------------|---------------|-------------|
| Query Match           | 45.4%;       | Score 69;       | DB 4;         | Length 768; |
| Best Local Similarity | 48.6%;       | Pred. No. 0.12; |               |             |
| Matches 17;           | Conservative | 1;              | Mismatches 3; | Indels 14;  |
|                       |              |                 |               | Gaps 2;     |

Qy 1 AYCGFLR-----PGVSSENLSAVATGNXGCGAFG 29  
||| ||: || |||||  
Db 478 AYIGFVHMVTPPPG-----VATGNWCGCGAFG 504

Search completed: May 26, 2004, 18:49:31  
Job time : 1.92926 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 2.41959 Seconds  
(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-12  
Perfect score: 152  
Sequence: 1 AYCGLRPGVSSSENLSAVATGNKXGCGAFG 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues  
Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 150   | 98.7        | 29     | 9  | US-09-973-451-12     |
| 2          | 149   | 98.0        | 976    | 9  | US-09-973-451-4      |
| 3          | 149   | 98.0        | 977    | 9  | US-09-973-451-2      |
| 4          | 144   | 94.7        | 968    | 9  | US-09-973-451-6      |
| 5          | 69    | 45.4        | 768    | 9  | US-09-973-451-8      |
| 6          | 67    | 44.1        | 546    | 12 | US-10-425-114-60000  |
| 7          | 66    | 43.4        | 180    | 12 | US-10-424-599-156445 |
| 8          | 60.5  | 39.8        | 543    | 14 | US-10-331-061-10     |
| 9          | 55.5  | 36.5        | 768    | 15 | US-10-369-493-2214   |
| 10         | 53    | 34.9        | 75     | 9  | US-09-993-844-46     |
| 11         | 53    | 34.9        | 75     | 16 | US-10-633-438-32     |
| 12         | 53    | 34.9        | 403    | 14 | US-10-251-385-114    |
| 13         | 53    | 34.9        | 403    | 14 | US-10-251-385-224    |
| 14         | 53    | 34.9        | 403    | 14 | US-10-225-567A-540   |
| 15         | 53    | 34.9        | 403    | 14 | US-10-290-078-18     |

|    |      |      |      |    |                      |                    |
|----|------|------|------|----|----------------------|--------------------|
| 16 | 53   | 34.9 | 403  | 15 | US-10-353-690-10     | Sequence 10, Appl  |
| 17 | 53   | 34.9 | 426  | 12 | US-10-311-671-1      | Sequence 1, Appl   |
| 18 | 53   | 34.9 | 445  | 15 | US-10-240-145-53     | Sequence 53, Appl  |
| 19 | 53   | 34.9 | 445  | 15 | US-10-240-145-139    | Sequence 139, Appl |
| 20 | 53   | 34.9 | 2762 | 12 | US-10-142-426-13     | Sequence 13, Appl  |
| 21 | 53   | 34.9 | 2762 | 14 | US-10-123-155-13     | Sequence 13, Appl  |
| 22 | 53   | 34.9 | 2762 | 14 | US-10-146-731-13     | Sequence 13, Appl  |
| 23 | 53   | 34.9 | 2762 | 14 | US-10-140-472-13     | Sequence 13, Appl  |
| 24 | 53   | 34.9 | 2762 | 14 | US-10-141-761-13     | Sequence 13, Appl  |
| 25 | 53   | 34.9 | 2762 | 14 | US-10-142-885-13     | Sequence 13, Appl  |
| 26 | 53   | 34.9 | 2762 | 14 | US-10-158-790-13     | Sequence 13, Appl  |
| 27 | 53   | 34.9 | 2762 | 15 | US-10-137-871-13     | Sequence 13, Appl  |
| 28 | 53   | 34.9 | 2762 | 15 | US-10-140-923-13     | Sequence 13, Appl  |
| 29 | 53   | 34.9 | 2762 | 15 | US-10-141-756-13     | Sequence 13, Appl  |
| 30 | 53   | 34.9 | 2762 | 15 | US-10-141-759-13     | Sequence 13, Appl  |
| 31 | 53   | 34.9 | 2762 | 15 | US-10-140-805-13     | Sequence 13, Appl  |
| 32 | 53   | 34.9 | 2762 | 15 | US-10-140-864-13     | Sequence 13, Appl  |
| 33 | 52   | 34.2 | 181  | 12 | US-10-424-599-17700  | Sequence 17700, A  |
| 34 | 51   | 33.6 | 371  | 15 | US-10-369-493-19533  | Sequence 19533, A  |
| 35 | 51   | 33.6 | 379  | 9  | US-09-738-626-5142   | Sequence 5142, Ap  |
| 36 | 50.5 | 33.2 | 501  | 12 | US-10-282-122A-63359 | Sequence 63359, A  |
| 37 | 50   | 32.9 | 372  | 12 | US-10-424-599-269980 | Sequence 269980,   |
| 38 | 49.5 | 32.6 | 726  | 9  | US-09-973-451-10     | Sequence 10, Appl  |
| 39 | 49   | 32.2 | 559  | 14 | US-10-271-697-4      | Sequence 4, Appl   |
| 40 | 49   | 32.2 | 568  | 9  | US-09-815-242-5140   | Sequence 5140, Ap  |
| 41 | 49   | 32.2 | 568  | 12 | US-10-282-122A-43547 | Sequence 43547, A  |
| 42 | 49   | 32.2 | 575  | 12 | US-10-425-114-45940  | Sequence 45940, A  |
| 43 | 49   | 32.2 | 616  | 10 | US-09-946-374-24     | Sequence 24, Appl  |
| 44 | 49   | 32.2 | 616  | 12 | US-10-206-915-158    | Sequence 158, Appl |
| 45 | 49   | 32.2 | 616  | 12 | US-10-199-670-158    | Sequence 158, Appl |

ALIGNMENTS

RESULT 1  
US-09-973-451-12  
; Sequence 12, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 12  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-12  
Query Match 98.7%; Score 150; DB 9; Length 29;  
Best Local Similarity 100.0%; Pred. No. 1.6e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKXGCGAFG 29  
Db 1 AYCGLRPGVSSSENLSAVATGNKXGCGAFG 29

RESULT 2

US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4

Query Match 98.0%; Score 149; DB 9; Length 976;  
Best Local Similarity 96.6%; Pred. No. 1.1e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
|||  
Db 848 AYCGLRPGVSSSENLSAVATGNWCGGAFG 876  
|||

RESULT 3  
US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

Query Match 98.0%; Score 149; DB 9; Length 977;  
Best Local Similarity 96.6%; Pred. No. 1.1e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
|||  
Db 849 AYCGLRPGVSSSENLSAVATGNWCGGAFG 877  
|||

RESULT 4  
US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

Query Match 94.7%; Score 144; DB 9; Length 968;  
Best Local Similarity 93.1%; Pred. No. 6e-12;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
|||  
Db 840 AYCGLRPGVPSSENLSAVATGNWCGGAFG 868  
|||

RESULT 5  
US-09-973-451-8  
; Sequence 8, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-973-451-8

Query Match 45.4%; Score 69; DB 9; Length 768;  
Best Local Similarity 48.6%; Pred. No. 0.31;  
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY 1 AYCGLR-----PGVSSSENLSAVATGNXGCGAFG 29  
|||  
Db 478 AYIGFVHMVTPPPG-----VATGNWCGGAFG 504  
|||

```
RESULT 6
US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-262-E12_FLI.pep
US-10-425-114-60000

Query Match 44.1%; Score 67; DB 12; Length 546;
Best Local Similarity 30.6%; Pred. No. 0.42;
Matches 15; Conservative 4; Mismatches 10; Indels 20; Gaps 1;

QY 1 AYCGFLR-----PGVSSSENLSAVATGNXGCGAFG 29
|:|
|:|
|:|
Db 407 AFCGFFDQSKHLYAKLFQDLHNKDDFSSINSSEYVGVSIGNWGGCGAFG 455

RESULT 7
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445

Query Match 43.4%; Score 66; DB 12; Length 180;
Best Local Similarity 61.9%; Pred. No. 0.17;
Matches 13; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 9 GVSSENLSAVATGNXGCGAFG 29
|:|
|:|
|:|
Db 67 GMDQNNIGVATGNWGGCGAFG 87

RESULT 8
US-10-331-061-10
; Sequence 10, Application US/10331061
; Publication No. US20030101486A1
; GENERAL INFORMATION:
; APPLICANT: Lassner, Michael
```

```
; APPLICANT: Metz, James G
; APPLICANT: Facciotti, Daniel
; TITLE OF INVENTION: SCHIZOCHYTRIUM PKS GENES
; FILE REFERENCE: CGNE.131.02US
; CURRENT APPLICATION NUMBER: US/10/331,061
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: US/09/231,899
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/048,650
; PRIOR FILING DATE: 1997-06-04
; PRIOR APPLICATION NUMBER: 09/090,793
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
; US-10-331-061-10

Query Match 39.8%; Score 60.5; DB 14; Length 543;
Best Local Similarity 46.7%; Pred. No. 3.6;
Matches 14; Conservative 4; Mismatches 9; Indels 3; Gaps 2;

QY 2 YCGFLRPGVSSSENLSAVATGNXG--CGAFG 29
|:|
|:|
|:|
Db 99 YAGAMANGISSEEL-VIALGQAGILCGSFG 127

RESULT 9
US-10-369-493-2214
; Sequence 2214, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2214
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(768)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-2214

Query Match 36.5%; Score 55.5; DB 15; Length 768;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 14; Conservative 1; Mismatches 8; Indels 3; Gaps 1;

QY 1 AYCGFLRPGVSSSENLSAVATGNXGCG 26
|:|
|:|
|:|
Db 498 ALCDF---GLSKANLSANATTNFCG 520

RESULT 10
US-09-993-844-46
; Sequence 46, Application US/09993844
; Patent No. US20020106739A1
; GENERAL INFORMATION:
; APPLICANT: Oakley, Robert H.
; APPLICANT: Barak, Lawrence S.
```



Thu May 27 09:55:56 2004

us-09-302-812-12.rapb

APPLICANT: Laporte, Stephane A.  
APPLICANT: Caron, Marc G.  
TITLE OF INVENTION: Modified G-Protein Coupled Receptors  
FILE REFERENCE: 033072-026  
CURRENT APPLICATION NUMBER: US/09/993,844  
CURRENT FILING DATE: 2001-11-05  
PRIOR APPLICATION NUMBER: US 60/245,772  
PRIOR FILING DATE: 2000-11-03  
PRIOR APPLICATION NUMBER: US 60/260,363  
PRIOR FILING DATE: 2001-01-08  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 46  
LENGTH: 75  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-993-844-46

Query Match 34.9%; Score 53; DB 9; Length 75;  
Best Local Similarity 45.2%; Pred. No. 4.8;  
Matches 14; Conservative 4; Mismatches 11; Indels 11; Gaps 1;

QY 1 AYCGLRPGVSSSENLSAVATGNXGC--GAFG 29  
Db 24 ACCHLRPRHSSHLSRMTTGTSLCDVGS LG 54

RESULT 11  
US-10-633-438-32  
Sequence 32, Application US/10633438  
Publication No. US20040091946A1  
GENERAL INFORMATION:  
APPLICANT: Oakley, Robert H.  
APPLICANT: Barak, Lawrence S.  
APPLICANT: Laporte, Stephane A.  
APPLICANT: Caron, Marc G.  
TITLE OF INVENTION: Methods of Screening Compositions for G Protein-Coupled Receptor  
TITLE OF INVENTION: Desensitization Inhibitory Activity  
FILE REFERENCE: 033072-044  
CURRENT APPLICATION NUMBER: US/10/633,438  
CURRENT FILING DATE: 2003-08-01  
PRIOR APPLICATION NUMBER: US 09/993,844  
PRIOR FILING DATE: 2001-11-05  
PRIOR APPLICATION NUMBER: US 60/245,772  
PRIOR FILING DATE: 2000-11-03  
PRIOR APPLICATION NUMBER: US 60/260,363  
PRIOR FILING DATE: 2001-01-08  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 32  
LENGTH: 75  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-633-438-32

Query Match 34.9%; Score 53; DB 16; Length 75;  
Best Local Similarity 45.2%; Pred. No. 4.8;  
Matches 14; Conservative 4; Mismatches 11; Indels 11; Gaps 1;

QY 1 AYCGLRPGVSSSENLSAVATGNXGC--GAFG 29  
Db 24 ACCHLRPRHSSHLSRMTTGTSLCDVGS LG 54

RESULT 12  
US-10-251-385-114  
Sequence 114, Application US/10251385  
Publication No. US20030105292A1  
GENERAL INFORMATION:  
APPLICANT: Behan, Dominic P.  
APPLICANT: Chalmers, Derek T.  
APPLICANT: Liaw, Chen W.  
TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G

TITLE OF INVENTION: Protein-Coupled  
TITLE OF INVENTION: Receptors  
FILE REFERENCE: AREN-0040  
CURRENT APPLICATION NUMBER: US/10/251,385  
CURRENT FILING DATE: 2002-09-20  
PRIOR APPLICATION NUMBER: US/09/170,496  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 294  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 114  
LENGTH: 403  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-251-385-114

Query Match 34.9%; Score 53; DB 14; Length 403;  
Best Local Similarity 45.2%; Pred. No. 31;  
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 AYCGLRPGVSSSENLSAVATGNXGC--GAFG 29  
Db 352 ACCHLRPRHSSHLSRMTTGTSLCDVGS LG 382

RESULT 13  
US-10-251-385-224  
Sequence 224, Application US/10251385  
Publication No. US20030105292A1  
GENERAL INFORMATION:  
APPLICANT: Behan, Dominic P.  
APPLICANT: Chalmers, Derek T.  
APPLICANT: Liaw, Chen W.  
TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G  
TITLE OF INVENTION: Protein-Coupled  
TITLE OF INVENTION: Receptors  
FILE REFERENCE: AREN-0040  
CURRENT APPLICATION NUMBER: US/10/251,385  
CURRENT FILING DATE: 2002-09-20  
PRIOR APPLICATION NUMBER: US/09/170,496  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 294  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 224  
LENGTH: 403  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-251-385-224

Query Match 34.9%; Score 53; DB 14; Length 403;  
Best Local Similarity 45.2%; Pred. No. 31;  
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 AYCGLRPGVSSSENLSAVATGNXGC--GAFG 29  
Db 352 ACCHLRPRHSSHLSRMTTGTSLCDVGS LG 382

RESULT 14  
US-10-225-567A-540  
Sequence 540, Application US/10225567A  
Publication No. US20030113798A1  
GENERAL INFORMATION:  
APPLICANT: LifeSpan Biosciences  
APPLICANT: Brown, Joseph P.  
APPLICANT: Burmer, Glenna C.  
APPLICANT: Roush, Christine L.  
TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 1920-4-4  
CURRENT APPLICATION NUMBER: US/10/225,567A  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/257,144  
PRIOR FILING DATE: 2000-12-19  
NUMBER OF SEQ ID NOS: 2292

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 540
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-540

Query Match      34.9%; Score 53; DB 14; Length 403;
Best Local Similarity 45.2%; Pred. No. 31;
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

Qy      1 AYCGFLRPGVSSSENLSAVATGNXGC--GAFG 29
      | | | | | | | | | | : | | : | | : |
Db      352 ACCHRLRPRHSSHLSLRTTGTSLCDVGS LG 382

RESULT 15
US-10-290-078-18
; Sequence 18, Application US/10290078
; Publication No. US20030124596A1
; GENERAL INFORMATION:
; APPLICANT: Carroll, Joseph A.
; TITLE OF INVENTION: Methods and Compositions for Treating
; TITLE OF INVENTION: Hematological Disorders Using 232, 2059, 10630, 12848, 13875,
; TITLE OF INVENTION: 14395, 14618, 17692 or 58874
; FILE REFERENCE: MPI2001-288P1(M)
; CURRENT APPLICATION NUMBER: US/10/290,078
; CURRENT FILING DATE: 2002-11-07
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-290-078-18

Query Match      34.9%; Score 53; DB 14; Length 403;
Best Local Similarity 45.2%; Pred. No. 31;
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

Qy      1 AYCGFLRPGVSSSENLSAVATGNXGC--GAFG 29
      | | | | | | | | | | : | | : | | : |
Db      352 ACCHRLRPRHSSHLSLRTTGTSLCDVGS LG 382

Search completed: May 26, 2004, 19:19:01
Job time : 3.41959 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 0.797763 Seconds  
(without alignments)  
3496.721 Million cell updates/sec

Title: US-09-302-812-12  
Perfect score: 152  
Sequence: 1 AYGGLRPGVSSNLSAVATGNXGCGAFG 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 60.5  | 39.8        | 543    | 2 T30186 | hypothetical prote |
| 2          | 60    | 39.5        | 997    | 2 B84726 | probable poly(ADP- |
| 3          | 55.5  | 36.5        | 696    | 2 S55694 | protein kinase (EC |
| 4          | 51    | 33.6        | 511    | 2 S10527 | endoglucanase B pr |
| 5          | 51    | 33.6        | 3716   | 2 E70969 | probable PPE prote |
| 6          | 50    | 32.9        | 282    | 2 A97214 | uncharacterized co |
| 7          | 49.5  | 32.6        | 504    | 2 E75615 | conserved hypothet |
| 8          | 49.5  | 32.6        | 726    | 2 T21138 | hypothetical prote |
| 9          | 49    | 32.2        | 568    | 2 D83182 | hypothetical prote |
| 10         | 48.5  | 31.9        | 516    | 2 D82279 | sodium/alanine sym |
| 11         | 48    | 31.6        | 433    | 2 AE2658 | glycolate oxidase  |
| 12         | 48    | 31.6        | 433    | 2 B97440 | glycolate oxidase  |
| 13         | 48    | 31.6        | 442    | 2 S77484 | preprotein translo |
| 14         | 47.5  | 31.2        | 411    | 2 T15705 | hypothetical prote |
| 15         | 47    | 30.9        | 129    | 2 AB2501 | hypothetical prote |
| 16         | 47    | 30.9        | 538    | 2 T07740 | probable inorganic |
| 17         | 47    | 30.9        | 538    | 2 T05714 | probable inorganic |
| 18         | 47    | 30.9        | 538    | 2 T07164 | probable inorganic |
| 19         | 47    | 30.9        | 595    | 2 A38628 | threonine ammonia- |
| 20         | 47    | 30.9        | 823    | 2 D86165 | protein F15K9.3 [i |
| 21         | 46.5  | 30.6        | 321    | 2 T45053 | hypothetical prote |
| 22         | 46.5  | 30.6        | 379    | 2 T52405 | hypothetical prote |
| 23         | 46.5  | 30.6        | 394    | 2 C81333 | probable efflux pu |
| 24         | 46.5  | 30.6        | 413    | 2 B44285 | methyiaspartate am |
| 25         | 46    | 30.3        | 186    | 2 F96718 | hypothetical prote |
| 26         | 46    | 30.3        | 262    | 2 B70780 | hypothetical prote |
| 27         | 46    | 30.3        | 273    | 2 S64054 | hypothetical prote |
| 28         | 46    | 30.3        | 542    | 2 T07808 | probable inorganic |
| 29         | 46    | 30.3        | 958    | 2 A82583 | conserved hypothet |

|    |      |      |      |          |                    |
|----|------|------|------|----------|--------------------|
| 30 | 46   | 30.3 | 2383 | 2 D64962 | probable membrane  |
| 31 | 46   | 30.3 | 2555 | 2 A40043 | notch protein homo |
| 32 | 45.5 | 29.9 | 424  | 2 B56144 | aspartate carbamoy |
| 33 | 45.5 | 29.9 | 481  | 2 F97472 | hypothetical prote |
| 34 | 45   | 29.6 | 96   | 2 FC4212 | gastrolith matrix  |
| 35 | 45   | 29.6 | 304  | 2 T38957 | probable GPR/FUN34 |
| 36 | 45   | 29.6 | 330  | 2 A40855 | homeotic protein H |
| 37 | 45   | 29.6 | 359  | 2 PQ0468 | threonine ammonia- |
| 38 | 45   | 29.6 | 359  | 2 F81036 | Mrp/NBP35 family p |
| 39 | 45   | 29.6 | 375  | 2 G81980 | hypothetical prote |
| 40 | 45   | 29.6 | 386  | 2 D83284 | probable acyl-CoA  |
| 41 | 45   | 29.6 | 405  | 4 A61181 | homeotic protein H |
| 42 | 45   | 29.6 | 412  | 2 T47142 | hypothetical prote |
| 43 | 45   | 29.6 | 474  | 2 A26421 | shufflon A - Esche |
| 44 | 45   | 29.6 | 624  | 2 T33868 | hypothetical prote |
| 45 | 45   | 29.6 | 882  | 2 A96830 | probable helicase, |

ALIGNMENTS

RESULT 1

T30186  
hypothetical protein 8 - Shewanella sp.  
C;Species: Shewanella sp.  
C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C;Accession: T30186  
R;Takeyama, H.; Takeda, D.; Yazawa, K.; Yamada, A.; Matsunaga, T.  
Microbiology 143, 2725-2731, 1997  
A;Title: Expression of the eicosapentaenoic acid synthesis gene cluster from Shewanella  
A;Reference number: Z20764; MUID:97419510; PMID:9274025  
A;Accession: T30186  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-543 <TAK>  
A;Cross-references: EMBL:U73935; NID:g2529413; PID:g2529421; PIDN:AAB81126.1  
A;Experimental source: strain SCRC-2738

Query Match 39.8%; Score 60.5; DB 2; Length 543;  
Best Local Similarity 46.7%; Pred. No. 0.8;  
Matches 14; Conservative 4; Mismatches 9; Indels 3; Gaps 2;

Qy 2 YCGFLRPGVSSNLSAVATGNXG--CGAFG 29

Db 99 YAGAMANGISSEEL-VIALGQAGILCGSFG 127

RESULT 2

B84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84726  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-997 <STO>  
A;Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g31870  
A;Map position: 2

Query Match 39.5%; Score 60; DB 2; Length 997;  
Best Local Similarity 51.9%; Pred. No. 1.8;  
Matches 14; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

Qy 5 FLRPGVSSNLS--SAVATGNXGCGAFG 29

Db 418 FYVEGVNEDHEDDGVATGNWGGVFG 444

RESULT 3

S55694

protein kinase (EC 2.7.1.37) sck1, CAMP-dependent - fission yeast (Schizosaccharomyces p

C;Species: Schizosaccharomyces pombe

C;Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 17-Nov-2000

C;Accession: S55694; T38040

R;Jin, M.; Fujita, M.; Culley, B.M.; Apolinario, E.; Yamamoto, M.; Maundrell, K.; Hoffma

Genetics 140, 457-467, 1995

A;Title: sck1, a high copy number suppressor of defects in the CAMP-dependent protein ki

A;Reference number: S55694; MUID:96120227; PMID:7498728

A;Accession: S55694

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-696 <JIN>

A;Cross-references: GB:D38108; NID:g1136301; PIDN:BAA07286.1; PID:d1007864; PID:g1136302

A;Note: the authors translated the codon GAT for residue 687 as His

R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Whitehead, S.; Churcher, C.M.

submitted to the EMBL Data Library, August 1999

A;Reference number: Z21764

A;Accession: T38040

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-176,'F',178-198,'A',200-696 <MCD>

A;Cross-references: EMBL:AL109951; PIDN:CAB53053.1; GSPDB:GN00066; SPDB:SPAC1B9.02C

A;Experimental source: strain 972h-; cosmid c1B9

C;Genetics:

A;Gene: sck1; SPDB:SPAC1B9.02C

A;Map position: 1

A;Introns: 80/3; 311/1; 633/2

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C;Keywords: ATP; phosphotransferase; protein kinase

F;300-563/Domain: protein kinase homology <KIN>

F;309-316/Region: protein kinase ATP-binding motif

Query Match 36.5%; Score 55.5; DB 2; Length 696;

Best Local Similarity 53.8%; Pred. No. 5.5;

Matches 14; Conservative 1; Mismatches 8; Indels 3; Gaps 1;

QY 1 AYCGFLRPGVSSSENLSAVATGNXGCG 26

Db 443 ALCDF---GLSKANLSANATTNTFCG 465

RESULT 4

S10527

endoglucanase B precursor - Pseudomonas fluorescens

C;Species: Pseudomonas fluorescens

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999

C;Accession: S10527

R;Gilbert, H.J.; Hall, J.; Hazlewood, G.P.; Ferreira, L.M.A.

Mol. Microbiol. 4, 759-767, 1990

A;Title: The N-terminal region of an endoglucanase from Pseudomonas fluorescens subspeci

A;Reference number: S10527; MUID:90355836; PMID:2117693

A;Accession: S10527

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-511 <GIL>

A;Cross-references: EMBL:X52615; NID:g45497; PIDN:CAA36844.1; PID:g45498

C;Superfamily: bacterial cellulose-binding domain homology; glycosidase GWGW domain homo

F;31-128/Domain: bacterial cellulose-binding domain homology <HCB>

F;180-217/Domain: glycosidase GWGW domain homology <GWG>

F;32-127/Disulfide bonds: #status predicted

Query Match 33.6%; Score 51; DB 2; Length 511;

Best Local Similarity 45.5%; Pred. No. 18;

Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 8 PGVSSSENLSAVATGNXGCGAFG 29

Db 164 PGTSSSSSSSVLTGAQACNWWYG 185

RESULT 5

E70969

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C;Accession: E70969

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: E70969

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-3716 <COL>

A;Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15735.1; PID:e120226

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: PPE

Query Match 33.6%; Score 51; DB 2; Length 3716;

Best Local Similarity 42.3%; Pred. No. 1.4e+02;

Matches 11; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 4 GFLRPGVSSSENLSAVATGNXGCGAFG 29

Db 2638 GFANQGVNIGLANTGTGNIIGLVG 2663

RESULT 6

A97214

uncharacterized conserved protein CAC2549 [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C;Accession: A97214

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: A97214

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-282 <KUR>

A;Cross-references: GB:AE001437; PIDN:AAK80500.1; PID:g15025572; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC2549

C;Superfamily: Deinococcus radiodurans hypothetical protein DRB0099

Query Match 32.9%; Score 50; DB 2; Length 282;

Best Local Similarity 50.0%; Pred. No. 14;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 12 SENLSAVATGNXGCGAFG 29

Db 213 SKNPKAIVLGAFGCGVFG 230

RESULT 7

E75615

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C;Accession: E75615

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 0.464631 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-12  
Perfect score: 152  
Sequence: 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID        | Description         |
|------------|-------|-------------|--------|--------------|---------------------|
| 1          | 55.5  | 36.5        | 696    | 1 SCK1 SCHPO | P50530 schizosacch  |
| 2          | 51    | 33.6        | 511    | 1 GUNB PSEFL | P18126 pseudomonas  |
| 3          | 48.5  | 31.9        | 886    | 1 SM6B MOUSE | O54951 mus musculus |
| 4          | 48.5  | 31.9        | 887    | 1 SM6B RAT   | O70141 rattus norv  |
| 5          | 48    | 31.6        | 442    | 1 SECY_SYN3  | P77964 synechocyst  |
| 6          | 47.5  | 31.2        | 175    | 1 NU6M DUGDU | Q8w9m5 dugong dugo  |
| 7          | 47    | 30.9        | 595    | 1 THD1 LYCES | P25306 lycopersico  |
| 8          | 46.5  | 30.6        | 307    | 1 IMP3_HUMAN | Q96f15 homo sapien  |
| 9          | 46.5  | 30.6        | 413    | 1 MAAL CLOTT | Q05514 clostridium  |
| 10         | 46    | 30.3        | 262    | 1 Y877_MYCTU | Q10539 mycobacteri  |
| 11         | 46    | 30.3        | 273    | 1 YGF0 YEAST | P53177 saccharomyc  |
| 12         | 46    | 30.3        | 1305   | 1 CYA9_XENLA | P98999 xenopus lae  |
| 13         | 46    | 30.3        | 2358   | 1 YEEJ_ECOLI | P76347 escherichia  |
| 14         | 46    | 30.3        | 2556   | 1 NTC1_HUMAN | P46531 homo sapien  |
| 15         | 45.5  | 29.9        | 424    | 1 PYRX_PSEPU | Q59712 pseudomonas  |
| 16         | 45    | 29.6        | 330    | 1 TLX1_HUMAN | P31314 homo sapien  |
| 17         | 45    | 29.6        | 359    | 1 THD1_SOLTU | P31212 solanum tub  |
| 18         | 45    | 29.6        | 474    | 1 SHU1_ECOLI | P09745 escherichia  |
| 19         | 44    | 28.9        | 211    | 1 RHOF_HUMAN | Q9hbh0 homo sapien  |
| 20         | 44    | 28.9        | 331    | 1 PME_ASPTU  | P17872 aspergillus  |
| 21         | 44    | 28.9        | 338    | 1 RTCA_ECOLI | P46849 escherichia  |
| 22         | 44    | 28.9        | 339    | 1 RTCA_SALTY | Q8zli0 salmonella   |
| 23         | 44    | 28.9        | 341    | 1 RTCA_PSEAE | Q9hvj9 pseudomonas  |
| 24         | 44    | 28.9        | 342    | 1 RTCA_ECO57 | P58127 escherichia  |
| 25         | 44    | 28.9        | 358    | 1 DPNF_ORYSA | Q40639 oryza sativ  |
| 26         | 44    | 28.9        | 360    | 1 FADH_ARYME | P80094 amycolatops  |
| 27         | 44    | 28.9        | 481    | 1 LEU2_COREF | Q8fpr3 corynebacte  |
| 28         | 44    | 28.9        | 481    | 1 LEU2_CORGL | P58946 corynebacte  |
| 29         | 44    | 28.9        | 485    | 1 LEU2_ACTT1 | Q44427 actinoplan   |
| 30         | 44    | 28.9        | 614    | 1 ASNO_BACSU | O05272 bacillus su  |
| 31         | 44    | 28.9        | 735    | 1 AD02_CAVPO | Q60411 cavila porce |
| 32         | 44    | 28.9        | 747    | 1 SPD1_NEPCL | P19837 nephila cla  |
| 33         | 44    | 28.9        | 971    | 1 AREA_GIBFU | P78688 gibberella   |

|    |      |      |      |              |                    |
|----|------|------|------|--------------|--------------------|
| 34 | 44   | 28.9 | 1093 | 1 AF17_HUMAN | P55198 homo sapien |
| 35 | 44   | 28.9 | 1365 | 1 SUZ2_DROME | P25172 drosophila  |
| 36 | 44   | 28.9 | 1801 | 1 LMB2_RAT   | P15800 rattus norv |
| 37 | 43.5 | 28.6 | 260  | 1 GLO2_MOUSE | Q99kb8 mus musculu |
| 38 | 43.5 | 28.6 | 260  | 1 GLO2_RAT   | O35952 rattus norv |
| 39 | 43.5 | 28.6 | 384  | 1 AGAS_ECOLI | P42907 escherichia |
| 40 | 43.5 | 28.6 | 408  | 1 FXDL_HUMAN | Q9nu39 homo sapien |
| 41 | 43.5 | 28.6 | 515  | 1 RPB2_METTW | P09844 methanobact |
| 42 | 43.5 | 28.6 | 842  | 1 VGLH_HSVBC | P27599 bovine herp |
| 43 | 43.5 | 28.6 | 1115 | 1 TBC2_CHLRE | Q8vxp3 chlamydomon |
| 44 | 43   | 28.3 | 115  | 1 PT22_STYPL | P28214 styela plic |
| 45 | 43   | 28.3 | 136  | 1 SC3_SCHCO  | P16933 schizophyll |

ALIGNMENTS

RESULT 1  
SCK1\_SCHPO  
ID SCK1 SCHPO STANDARD; PRT; 696 AA.  
AC P50530; Q9UTF3;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Serine/threonine-protein kinase sck1 (EC 2.7.1.37).  
GN SCK1 OR SPAC1B9.02C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96120227; PubMed=7493728;  
RA Jin M., Fujita M., Culley B., Apolinario E., Yamamoto M.,  
RA Maundrell K., Hoffman C.;  
RT "sck1, a high copy number suppressor of defects in the CAMP-dependent  
RT protein kinase pathway in fission yeast, encodes a protein homologous  
RT to the Saccharomyces cerevisiae SCH9 kinase.";  
RL Genetics 140:457-467(1995).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckie E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Usseery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. CAMP  
CC subfamily.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; D38108; BAA07286.1; --  
EMBL; AL109951; CAB53053.1; --  
PIR; S55694; S55694.  
HSP; P05132; 1CTP.  
GeneDB Spombe; SPAC1B9.02c; --  
InterPro; IPR000008; C2.  
InterPro; IPR008973; C2\_CalB.  
InterPro; IPR000961; Pkinase\_C.  
InterPro; IPR000719; Prot\_kinase.  
InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
InterPro; IPR002290; Ser\_thr\_pkinase.  
InterPro; IPR001245; Tyr\_pkinase.  
Pfam; PF00168; C2; 1.  
Pfam; PF00069; pkinase; 1.  
Pfam; PF00433; pkinase\_C; 1.  
PRINTS; PR00109; TYRKINASE.  
ProDom; PD000001; Prot\_kinase; 1.  
SMART; SM00239; C2; 1.  
SMART; SM00133; S\_TK\_X; 1.  
SMART; SM00220; S\_TK; 1.  
PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
Transferrase; Serine/threonine-protein kinase; ATP-binding; cAMP.  
FT DOMAIN 302 563 PROTEIN\_KINASE.  
FT NP\_BIND 308 316 ATP (BY SIMILARITY).  
FT BINDING 331 331 ATP (BY SIMILARITY).  
FT ACT\_SITE 428 428 BY SIMILARITY.  
FT CONFLICT 199 199 A -> R (IN REF. 1).  
SQ SEQUENCE 696 AA; 78594 MW; A7B05F5EE4D42AF7 CRC64;

Query Match 36.5%; Score 55.5; DB 1; Length 696;  
Best Local Similarity 53.8%; Pred. No. 1.8;  
Matches 14; Conservative 1; Mismatches 8; Indels 3; Gaps 1;

QY 1 AYCGFLRPGVSSSENLSAVATGNXGCG 26  
Db 443 ALCDF---GLSKANLSANATNTNFCG 465

RESULT 2  
GUNB\_PSEFL  
ID GUNB\_PSEFL STANDARD; PRT; 511 AA.  
AC P18126;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase) (EGB).  
GN CELB.  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.  
RC STRAIN=Sp. Cellulosa;  
RX MEDLINE=90355836; PubMed=2117693;  
RA Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;  
RT "The N-terminal region of an endoglucanase from Pseudomonas  
RT fluorescens subspecies cellulosa constitutes a cellulose-binding  
RT domain that is distinct from the catalytic centre."  
RL Mol. Microbiol. 4:759-767(1990).  
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-

CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
CC GLUCANS. EGB IS MOST ACTIVE AGAINST BARLEY BETA-GLUCAN, BUT SHOWED  
CC SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRISTALLINE CELLULOSE.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)  
CC domain.  
CC -!- SIMILARITY: Belongs to cellulase family K (family 45 of glycosyl  
CC hydrolases).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X52615; CAA36844.1; --  
CC PIR; S10527; S10527.  
CC HSP; P43316; 2ENG.  
CC InterPro; IPR001919; Bac\_celose-bind.  
CC InterPro; IPR009009; Barwin\_like.  
CC InterPro; IPR009031; CBDX.  
CC InterPro; IPR008965; Cellul bind.  
CC InterPro; IPR002883; Dockerin CBD\_5.  
CC InterPro; IPR000334; Glyco\_hydro\_45.  
CC Pfam; PF02013; CBM\_10; 1.  
CC Pfam; PF00553; CBM\_2; 1.  
CC Pfam; PF02015; Glyco\_hydro\_45; 1.  
CC SMART; SM00637; CBD\_II; 1.  
CC PROSITE; PS00561; GLYCOSYL\_HYDROL\_F45; 1.  
CC PROSITE; PS01140; GLYCOLASE; Hydrolase; Glycosidase; Signal; Periplasmic.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Periplasmic.  
FT SIGNAL 1 29  
FT CHAIN 30 511 ENDOGLUCANASE B.  
FT DOMAIN 30 131 CELLULOSE-BINDING (BY SIMILARITY).  
FT DOMAIN 132 173 SER-RICH (LINKER).  
FT DOMAIN 223 259 SER-RICH.  
FT DISULFID 32 127 BY SIMILARITY.  
FT ACT\_SITE 276 276 NUCLEOPHILE (BY SIMILARITY).  
FT ACT\_SITE 393 393 PROTON DONOR (BY SIMILARITY).  
FT SEQUENCE 511 AA; 52078 MW; 3C3119D998291D8E CRC64;

Query Match 33.6%; Score 51; DB 1; Length 511;  
Best Local Similarity 45.5%; Pred. No. 6.1;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 8 PGVSSENLSAVATGNXGCGAFG 29  
Db 164 PGTSSSSSSSVLTGAQACNWTG 185

RESULT 3  
SM6B\_MOUSE  
ID SM6B\_MOUSE STANDARD; PRT; 886 AA.  
AC O54951;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Semaphorin 6B precursor (Semaphorin VIB) (Sema VIB) (Semaphorin N)  
DE (Sema N).  
GN SEMA6B OR SEMAN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98027184; PubMed=9361278;  
RA Eckhardt F., Behar O., Calautti E., Yonezawa K., Nishimoto I.,  
RA Fishman M.C.;



RT "A novel transmembrane semaphorin can bind c-src.";

RL Mol. Cell. Neurosci. 9:409-419(1997).

CC -!- SUBUNIT: HOMODIMER. BINDS SPECIFICALLY THE SH3 DOMAIN OF THE

CC PROTOONCOGENE C-SRC.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- TISSUE SPECIFICITY: During development it is expressed in

CC subregions of the nervous system and is particularly prominent in

CC muscle. In adulthood, it is expressed ubiquitously.

CC -!- SIMILARITY: Belongs to the semaphorin family.

CC -!- SIMILARITY: Contains 1 Sema domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF036585; AAC00493.1; -.

CC MGD; MGI:1202889; Sema6b.

CC InterPro; IPR003659; Plexin-like.

CC InterPro; IPR001627; Sema.

CC Pfam; PF01403; Sema; 1.

CC SMART; SM00423; PSI; 1.

CC SMART; SM00630; Sema; 1.

CC Signal; Transmembrane; Multigene family; Neurogenesis; Glycoprotein;

KW Developmental protein.

FT SIGNAL 1 26

FT CHAIN 27 886

FT DOMAIN 27 886

FT TRANSMEM 606 626

FT DOMAIN 627 886

FT DOMAIN 239 549

FT DOMAIN 751 754

FT CARBOHYD 75 75

FT CARBOHYD 156 156

FT CARBOHYD 292 292

FT CARBOHYD 387 387

FT CARBOHYD 442 442

FT CARBOHYD 463 463

FT CARBOHYD 886 AA; 95466 MW; E5F56D125CDA574D CRC64;

SQ SEQUENCE

Query Match 31.9%; Score 48.5; DB 1; Length 886;

Best Local Similarity 35.3%; Pred. No. 26;

Matches 12; Conservative 5; Mismatches 6; Indels 11; Gaps 2;

QY 2 YCG-----FLRPGVSS---ENLSAVATGNXG 24

Db 545 YCGWAPDGCIFLRPGTSATFEQDVSGASTSGLG 578

RESULT 4

SM6B RAT STANDARD; PRT; 887 AA.

AC O70141;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Semaphorin 6B precursor (Semaphorin Z) (Sema Z).

GN Sema6b.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Brain;

RX MEDLINE=98087397; PubMed=9427525;

RA Kikuchi K., Ishida H., Kimura T.;

RT "Molecular cloning of a novel member of semaphorin family genes,

RT semaphorin Z.";

RL Brain Res. Mol. Brain Res. 51:229-237(1997).

CC -!- FUNCTION: MAY PLAY A ROLE IN BOTH PERIPHERAL AND CENTRAL NERVOUS

CC SYSTEM DEVELOPMENT.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- DEVELOPMENTAL STAGE: DETECTED IN THE FIRST BRANCHIAL ARCH OF

CC EMBRYONIC DAY 11 (E11) EMBRYO, AND SUBSEQUENTLY IN THE MYOTOMES

CC AND THE DORSAL ROOT GANGLIA IN DEVELOPING SOMITES FROM E11.5

CC THROUGH E13.5, BUT NOT IN THE BRAIN. HOWEVER, AT E15, 18, 21 AND

CC P0, SEMAZ WAS HIGHLY EXPRESSED IN THE BRAIN.

CC -!- SIMILARITY: Belongs to the semaphorin family.

CC -!- SIMILARITY: Contains 1 Sema domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AB000776; BAA25687.1; -.

CC InterPro; IPR003659; Plexin-like.

CC InterPro; IPR001627; Sema.

CC Pfam; PF01403; Sema; 1.

CC SMART; SM00423; PSI; 1.

CC SMART; SM00630; Sema; 1.

CC Signal; Transmembrane; Multigene family; Neurogenesis; Glycoprotein;

KW Developmental protein.

FT SIGNAL 1 26

FT CHAIN 27 887

FT DOMAIN 27 605

FT TRANSMEM 606 626

FT DOMAIN 627 887

FT DOMAIN 239 549

FT CARBOHYD 75 75

FT CARBOHYD 156 156

FT CARBOHYD 168 168

FT CARBOHYD 292 292

FT CARBOHYD 387 387

FT CARBOHYD 442 442

FT CARBOHYD 463 463

FT CARBOHYD 887 AA; 95752 MW; 09543F3F202CD301 CRC64;

SQ SEQUENCE

Query Match 31.9%; Score 48.5; DB 1; Length 887;

Best Local Similarity 35.3%; Pred. No. 26;

Matches 12; Conservative 5; Mismatches 6; Indels 11; Gaps 2;

QY 2 YCG-----FLRPGVSS---ENLSAVATGNXG 24

Db 545 YCGWAPDGCIFLRPGTSATFEQDVSGASTSGLG 578

RESULT 5

SECY SYNY3 STANDARD; PRT; 442 AA.

AC P77964;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Preprotein translocase secy subunit.

GN SECY OR SLL1814.

OS Synechocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

OX NCBI\_TaxID=1148;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,

RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -|- FUNCTION: Involved in protein export. Interacts with secA and secE  
CC to allow the translocation of proteins across the plasma membrane,  
CC by forming part of a channel.  
CC -|- SUBUNIT: One of seven secretory proteins (secA-F and secY) that  
CC comprise the prokaryotic protein translocation apparatus.  
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -|- SIMILARITY: Belongs to the secY/SEC61-alpha family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D90905; BAA17331.1; -.  
DR PIR; S77484; S77484.  
DR InterPro; IPR002208; SecY.  
DR Pfam; PF00344; secY; 1.  
DR PRINTS; PR00303; SECYTRNLCASE.  
DR TIGRFAMs; TIGR00967; 3a0501s007; 1.  
DR PROSITE; PS00755; SECY\_1; 1.  
DR PROSITE; PS00756; SECY\_2; 1.  
KW Protein transport; Translocation; Transmembrane; Complete proteome.  
SQ SEQUENCE 442 AA; 47999 MW; 641D89CF9C0655E5 CRC64;  
  
Query Match 31.6%; Score 48; DB 1; Length 442;  
Best Local Similarity 56.2%; Pred. No. 15;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
  
QY 14 NLSAVATGNXGCGAFG 29  
:|:|||||  
DB 289 SLAGFATGNEGLGFG 304  
  
RESULT 6  
NU6M\_DUGDU STANDARD; PRT; 175 AA.  
ID NU6M\_DUGDU  
AC Q8W9M5;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).  
GN MTND6 OR ND6 OR NADH6.  
OS Dugong dugon (Dugong).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Sirenia; Dugongidae; Dugong.  
OX NCBI\_TaxID=29137;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22056029; PubMed=12034869;  
RA Arnason U., Adegoke J.A., Bodin K., Born E.W., Esa Y.B., Gullberg A.,  
RA Nilsson M., Short R.V., Xu X., Janke A.;  
RT "Mammalian mitogenomic relationships and the root of the eutherian  
tree.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:8151-8156(2002).  
CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ421723; CAD18919.1; -.  
DR InterPro; IPR001457; Oxidored\_q3.  
DR Pfam; PF00499; oxidored\_q3; 1.

KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
SQ SEQUENCE 175 AA; 18441 MW; 063F763F5FE614A3 CRC64;  
  
Query Match 31.2%; Score 47.5; DB 1; Length 175;  
Best Local Similarity 42.9%; Pred. No. 6.6;  
Matches 15; Conservative 4; Mismatches 5; Indels 11; Gaps 3;  
  
QY 4 GFLRPGVSSE-----NLSAVATGNXGCG---AFG 29  
||:|||||:|||||  
DB 15 GFV--GVSSKSPVPYGGGLGLVASGGVGCGIVVSFG 47  
  
RESULT 7  
THD1\_LYCES STANDARD; PRT; 595 AA.  
ID THD1\_LYCES  
AC P25306;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Threonine dehydratase biosynthetic, chloroplast precursor  
DE (EC 4.3.1.19) (Threonine deaminase) (TD).  
GN TD.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC Lamids; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 52-61.  
RC STRAIN=cv. Tiny Tim LA154;  
RX MEDLINE=91187855; PubMed=2011578;  
RA Samach A., Hareven D., Gutfinger T., Ken-Dror S., Lifschitz E.;  
RT "Biosynthetic threonine deaminase gene of tomato: isolation,  
RT structure, and upregulation in floral organs.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:2678-2682(1991).  
CC -|- CATALYTIC ACTIVITY: L-threonine = 2-oxobutanoate + NH(3).  
CC -|- COFACTOR: Pyridoxal phosphate.  
CC -|- ENZYME REGULATION: Allosterically inhibited by isoleucine.  
CC -|- PATHWAY: Isoleucine biosynthesis; first step.  
CC -|- SUBUNIT: Homotetramer.  
CC -|- SUBCELLULAR LOCATION: Chloroplast.  
CC -|- MISCELLANEOUS: Expression in mature flowers is increased over  
CC 500-fold.  
CC -|- SIMILARITY: Belongs to the serine/threonine dehydratase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M61914; AAA34171.1; -.  
DR EMBL; M61915; AAA68097.1; -.  
DR PIR; A38628; A38628.  
DR HSSP; P04968; 1TDJ.  
DR InterPro; IPR001926; B6 enzyme beta.  
DR InterPro; IPR000634; S/T\_dehydratse\_BS.  
DR InterPro; IPR005787; Thr\_dehydrataseI.  
DR InterPro; IPR001721; ThrDh\_C.  
DR Pfam; PF00291; PALP; 1.  
DR Pfam; PF00585; Thr\_dehydrat\_C; 2.  
DR TIGRFAMs; TIGR01124; ilva\_2Cterm; 1.  
DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
KW Isoleucine biosynthesis; Lyase; Chloroplast; Pyridoxal phosphate;  
KW Allosteric enzyme; Transist peptide.  
FT TRANSIT 1 51 CHLOROPLAST.  
FT CHAIN 52 595 THREONINE DEHYDRATASE BIOSYNTHETIC.  
FT BINDING 143 143 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 595 AA; 64937 MW; AC430BB5DD9F0348 CRC64;  
  
Query Match 30.9%; Score 47; DB 1; Length 595;









```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z72572; CAA96752.1; -.
CC PIR; S64054; S64054.
CC GermOnline; 141098; -.
CC SGD; S0003018; YGL050W.
CC InterPro; IPR003827; DUF207.
CC Pfam; PF02676; DUF207; 1.
CC KW Hypothetical protein.
CC SQ SEQUENCE 273 AA; 30806 MW; A36D61F625B0D8EF CRC64;

Query Match 30.3%; Score 46; DB 1; Length 273;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 CGFLRPGVSSSENLSAV 18
Db 156 CGFRESGIGSNNLVAI 171

RESULT 12
CYA9 XENLA STANDARD; PRT; 1305 AA.
ID CYA9 XENLA STANDARD; PRT; 1305 AA.
AC P98999;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenylate cyclase, type IX (EC 4.6.1.1) (ATP pyrophosphate-lyase)
DE (Adenylyl cyclase) (xIAC).
GN ADCY9.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=97228584; PubMed=9074644;
RA Torrejon M., Echeverria V., Retamales G., Herrera L., Hinrichs M.V.,
RA Olate J.;
RT "Molecular cloning and expression of an adenylyl cyclase from Xenopus
RT laevis oocytes."
RL FEBS Lett. 404:91-94(1997).
CC -!- FUNCTION: May play a fundamental role in situations where fine
CC interplay between intracellular calcium and cAMP determines the
CC cellular function. May be a physiologically relevant docking site
CC for calcineurin (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -!- COFACTOR: Binds 2 magnesium ions per subunit.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- DOMAIN: Composed of two homologous domains.
CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl
CC cyclase family.
CC -!- SIMILARITY: Contains 1 FKBP-type PPIase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46958; CAA87082.1; -.
CC HSSP; P26769; IAB8.
CC InterPro; IPR001054; G_cyclase.
CC Pfam; PF00211; guanylate_cyc; 2.
```

```
DR SMART; SM00044; CYCC; 2.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS0125; GUANYLATE_CYCLASES_2; 2.
KW Lyase; cAMP biosynthesis; Transmembrane; Glycoprotein; Repeat;
KW Metal-binding; Magnesium; Isomerase; Rotamase.
FT DOMAIN 1 110 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 111 131 POTENTIAL.
FT DOMAIN 132 134 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 135 155 POTENTIAL.
FT DOMAIN 156 164 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 165 185 POTENTIAL.
FT DOMAIN 186 209 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 210 229 POTENTIAL.
FT DOMAIN 230 235 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 236 253 POTENTIAL.
FT DOMAIN 254 274 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 275 295 POTENTIAL.
FT DOMAIN 296 778 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 779 799 POTENTIAL.
FT DOMAIN 800 810 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 811 831 POTENTIAL.
FT DOMAIN 832 859 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 860 880 POTENTIAL.
FT DOMAIN 881 904 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 905 911 POTENTIAL.
FT DOMAIN 912 932 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 933 966 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 967 987 POTENTIAL.
FT DOMAIN 988 1305 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1305 1305 PPIASE, FKBP-TYPE.
FT DOMAIN 152 157 POLY-PHE.
FT TRANSMEM 157 177 MAGNESIUM 1 AND 2 (BY SIMILARITY). (BY
FT METAL 177 393 MAGNESIUM 2 (VIA CARBONYL OXYGEN)
FT METAL 394 394 SIMILARITY).
FT METAL 437 437 MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 946 946 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 1305 AA; 145410 MW; 8E1A5B79466B556A CRC64;

Query Match 30.3%; Score 46; DB 1; Length 1305;
Best Local Similarity 38.7%; Pred. No. 93;
Matches 12; Conservative 6; Mismatches 5; Indels 8; Gaps 2;

QY 6 LRPGVSSSENLSAV---ATGNXGC---GAF 28
Db 188 LNPGLGSDLSNLTFSATGSSSCLSQVGSF 218

RESULT 13
YEEJ_ECOLI STANDARD; PRT; 2358 AA.
ID YEEJ_ECOLI STANDARD; PRT; 2358 AA.
AC P76347; P94750;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yeeJ.
GN YEEJ OR B1978.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
```





```
RC STRAIN=PPN-1;
RX MEDLINE=95204340; PubMed=7896697;
RA Schurr M.J., Vickrey J.F., Kumar A.P., Campbell A.L., Cunin R.,
RA Benjamin R.C., Shanley M.S., O'Donovan G.A.;
RT "Aspartate transcarbamoylase genes of pseudomonas putida: requirement
RT for an inactive dihydroorotase for assembly into the dodecameric
RT holoenzyme.";
RL J. Bacteriol. 177:1751-1759(1995).
CC -|- FUNCTION: NON-FUNCTIONAL DHOASE.
CC -|- SUBUNIT: HETERODECAMER OF 6 ACTIVE PYRB SUBUNITS AND 6 NON-
CC CATALYTIC PYRC' SUBUNITS.
CC -|- SIMILARITY: Belongs to the DHOase family. Pyrc' subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M97254; AAA69779.1; -.
DR PIR; B56144; B56144.
DR HAMAP; MF_00220; atypical; 1.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR005847; Pept_M38_regn.
DR Pfam; PF01979; Amidohydro_1; 1.
DR ProDom; PD000518; Urease; 1.
DR Pyrimidine biosynthesis.
KW SEQUENCE 424 AA; 44264 MW; 87159A625340AC7C CRC64;
SQ -----
Query Match 29.9%; Score 45.5; DB 1; Length 424;
Best Local Similarity 44.4%; Pred. No. 34;
Matches 12; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY 4 GFLRPGVSSNLS-AVATGNXGCGAFG 29
Db 125 GALTGLEGEQLAELVALRDTGCVAFG 151
```

Search completed: May 26, 2004, 18:41:19  
Job time : 2.46463 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 2.31439 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-12  
Perfect score: 152  
Sequence: 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_invertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriaph:\*
  - 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description         |
|------------|-------|-------------|--------|-----------|---------------------|
| 1          | 149   | 98.0        | 976    | 4 Q9Y4W7  | Q9Y4W7 homo sapien  |
| 2          | 149   | 98.0        | 976    | 4 Q86W56  | Q86W56 homo sapien  |
| 3          | 149   | 98.0        | 976    | 4 Q7Z742  | Q7Z742 homo sapien  |
| 4          | 149   | 98.0        | 977    | 6 O02776  | O02776 bos taurus   |
| 5          | 144   | 94.7        | 920    | 11 Q8CB72 | Q8CB72 mus musculus |
| 6          | 144   | 94.7        | 961    | 11 Q80YQ6 | Q80YQ6 mus musculus |
| 7          | 144   | 94.7        | 968    | 11 Q88622 | Q88622 mus musculus |
| 8          | 139   | 91.4        | 972    | 11 Q9QYM2 | Q9QYM2 rattus norv  |
| 9          | 69    | 45.4        | 723    | 5 Q960N8  | Q960N8 drosophila   |
| 10         | 69    | 45.4        | 768    | 5 O46043  | O46043 drosophila   |
| 11         | 60.5  | 39.8        | 543    | 2 O33907  | O33907 shewanella   |
| 12         | 60    | 39.5        | 548    | 10 Q9SKB3 | Q9SKB3 arabidopsis  |
| 13         | 57    | 37.5        | 522    | 10 Q8VYAL | Q8VYAL arabidopsis  |
| 14         | 55.5  | 36.5        | 513    | 16 Q8DF10 | Q8DF10 vibrio vuln  |
| 15         | 55    | 36.2        | 578    | 5 Q9VIS9  | Q9VIS9 drosophila   |
| 16         | 55    | 36.2        | 607    | 5 Q8MR68  | Q8MR68 drosophila   |

|    |      |      |      |           |                    |
|----|------|------|------|-----------|--------------------|
| 17 | 54   | 35.5 | 358  | 16 Q8G798 | Q8G798 bifidobacte |
| 18 | 53   | 34.9 | 426  | 4 Q8NE20  | Q8NE20 homo sapien |
| 19 | 53   | 34.9 | 426  | 4 Q9HB89  | Q9HB89 homo sapien |
| 20 | 52   | 34.2 | 485  | 16 Q92Q80 | Q92Q80 rhizobium m |
| 21 | 52   | 34.2 | 494  | 11 Q9D2R3 | Q9D2R3 mus musculu |
| 22 | 52   | 34.2 | 495  | 11 Q8BZQ2 | Q8BZQ2 mus musculu |
| 23 | 51.5 | 33.9 | 849  | 12 Q9WQY5 | Q9WQY5 bovine herp |
| 24 | 51   | 33.6 | 379  | 16 Q8NQE8 | Q8NQE8 corynebacte |
| 25 | 51   | 33.6 | 644  | 16 Q98CV7 | Q98CV7 rhizobium l |
| 26 | 51   | 33.6 | 3716 | 16 Q50379 | Q50379 mycobacteri |
| 27 | 50.5 | 33.2 | 1384 | 16 Q8EUY9 | Q8EUY9 mycoplasma  |
| 28 | 50   | 32.9 | 282  | 16 Q97G23 | Q97G23 clostridium |
| 29 | 50   | 32.9 | 618  | 13 Q8AVC4 | Q8AVC4 xenopus lae |
| 30 | 50   | 32.9 | 706  | 5 Q8SXG1  | Q8SXG1 drosophila  |
| 31 | 50   | 32.9 | 759  | 11 Q9JKC9 | Q9JKC9 rattus norv |
| 32 | 50   | 32.9 | 861  | 5 Q9VRT8  | Q9VRT8 drosophila  |
| 33 | 50   | 32.9 | 4019 | 5 Q9VJB4  | Q9VJB4 drosophila  |
| 34 | 49.5 | 32.6 | 307  | 16 Q88EZ3 | Q88EZ3 pseudomonas |
| 35 | 49.5 | 32.6 | 504  | 16 Q9RYW7 | Q9RYW7 deinococcus |
| 36 | 49.5 | 32.6 | 764  | 5 Q19637  | Q19637 caenorhabdi |
| 37 | 49.5 | 32.6 | 781  | 5 Q867X0  | Q867X0 caenorhabdi |
| 38 | 49   | 32.2 | 175  | 10 Q9FI91 | Q9FI91 arabidopsis |
| 39 | 49   | 32.2 | 214  | 16 Q88IV4 | Q88IV4 pseudomonas |
| 40 | 49   | 32.2 | 362  | 4 Q8TC55  | Q8TC55 homo sapien |
| 41 | 49   | 32.2 | 385  | 11 Q8BMW6 | Q8BMW6 mus musculu |
| 42 | 49   | 32.2 | 495  | 5 Q9GQ43  | Q9GQ43 giardia lam |
| 43 | 49   | 32.2 | 548  | 5 Q9GQ45  | Q9GQ45 giardia lam |
| 44 | 49   | 32.2 | 568  | 16 Q9HXS5 | Q9HXS5 pseudomonas |
| 45 | 49   | 32.2 | 815  | 10 Q9SWB4 | Q9SWB4 glycine max |

ALIGNMENTS

RESULT 1  
Q9Y4W7  
ID Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
AC Q9Y4W7;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9380098; PubMed=10449915;  
RA Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
RT hybridization.";  
RL CytoGenet. Cell Genet. 85:269-270 (1999).  
DR EMBL; AF005043; AAB61614.1; -.  
DR Genew; HGNC:8605; PARG.  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;

Query Match 98.0%; Score 149; DB 4; Length 976;  
Best Local Similarity 96.6%; Pred. No. 1e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876

```

RESULT 2
Q86W56
ID Q86W56 PRELIMINARY; PRT; 976 AA.
AC Q86W56;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to poly (ADP-ribose) glycohydrolase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050560; AAH50560.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;

Query Match 98.0%; Score 149; DB 4; Length 976;
Best Local Similarity 96.6%; Pred. No. 1e-12;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876

RESULT 3
Q7Z742
ID Q7Z742 PRELIMINARY; PRT; 976 AA.
AC Q7Z742;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876

RESULT 4
O02776
ID O02776 PRELIMINARY; PRT; 977 AA.
AC O02776;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN BPARG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97277328; PubMed=9115250;
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase."
RL J. Biol. Chem. 272:11895-11901 (1997).
DR EMBL; U78975; AAB53370.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;

Query Match 98.0%; Score 149; DB 6; Length 977;
Best Local Similarity 96.6%; Pred. No. 1e-12;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db 849 AYCGLRPGVSSSENLSAVATGNXGCGAFG 877

RESULT 5
Q8CB72
ID Q8CB72 PRELIMINARY; PRT; 920 AA.
AC Q8CB72;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
```

```
DR EMBL; AK036656; BAC29519.1; -.
DR MGD; MGI:1347094; PARG.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;
Query Match 94.7%; Score 144; DB 11; Length 920;
Best Local Similarity 93.1%; Pred. No. 5e-12;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db 841 AYCGLRPGVPSNLSAVATGNWCGGAFG 869

RESULT 6
Q80YQ6 PRELIMINARY; PRT; 961 AA.
AC Q80YQ6
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Poly (ADP-ribose) glycohydrolase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050892; AAH50892.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

Query Match 94.7%; Score 144; DB 11; Length 961;
Best Local Similarity 93.1%; Pred. No. 5.3e-12;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db 841 AYCGLRPGVPSNLSAVATGNWCGGAFG 869

RESULT 7
O88622 PRELIMINARY; PRT; 968 AA.
AC O88622
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly (ADP-ribose) glycohydrolase.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99380098; PubMed=10449915;
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
human chromosome 10q11.23 and mouse chromosome 14B by in situ
hybridization."
RL Cytogenet. Cell Genet. 85:269-270(1999).
DR EMBL; AF079557; AAC28735.1; -.
DR MGD; MGI:1347094; PARG.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
```

```
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;
Query Match 94.7%; Score 144; DB 11; Length 968;
Best Local Similarity 93.1%; Pred. No. 5.3e-12;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db 840 AYCGLRPGVPSNLSAVATGNWCGGAFG 868

RESULT 8
Q9QYM2 PRELIMINARY; PRT; 972 AA.
AC Q9QYM2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly (ADP-ribose) glycohydrolase.
GN PARG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BUF; TISSUE=Colon;
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,
RA Aoki Y., Nakgama H., Sugimura T.;
RL "Isolation and cloning of rat poly (ADP-ribose) glycohydrolase
(Parg).";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB019366; BAA87901.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;

Query Match 91.4%; Score 139; DB 11; Length 972;
Best Local Similarity 89.7%; Pred. No. 2.8e-11;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29
Db 844 AYCGLRPGVPSNLSAVATGNWCGGAFG 872

RESULT 9
Q960N8 PRELIMINARY; PRT; 723 AA.
AC Q960N8
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE LD42380p.
GN PARG OR EG:114E2.1 OR CG2864.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
```

DR EMBL; AY051955; AAK93379.1; -.  
DR FlyBase; FBgn0023216; PARG.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;  
  
Query Match 45.4%; Score 69; DB 5; Length 723;  
Best Local Similarity 48.6%; Pred. No. 0.27;  
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;  
  
QY 1 AYCGFLR-----PGVSSENLSAVATGNKXGCGAFG 29  
||| |||  
Db 433 AYIGFVHMVTPPPG-----VATGNWCGGAFG 459  
  
RESULT 10  
O46043 PRELIMINARY; PRT; 768 AA.  
AC O46043;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE PARG protein.  
GN PARG OR EG:114E2.1 OR CG2864.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M.D., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.

RA Ame J.-C., Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding Drosophila  
RT poly(ADP-ribose) glycohydrolase."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBBJ databases.  
RN [3]  
RP SEQUENCE OF 46-768 FROM N.A.  
RA Murphy L., Harris D., Barrell B.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBBJ databases.  
RN [4]  
RP SEQUENCE OF 46-768 FROM N.A.  
RA Benos P.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBBJ databases.  
DR EMBL; AB003428; AAF45886.1; -.  
DR EMBL; AF079556; AAC28734.1; -.  
DR EMBL; Z98254; CAB10913.1; -.  
DR FlyBase; FBgn0023216; PARG.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;  
  
Query Match 45.4%; Score 69; DB 5; Length 768;  
Best Local Similarity 48.6%; Pred. No. 0.29;  
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;  
  
QY 1 AYCGFLR-----PGVSSENLSAVATGNKXGCGAFG 29  
||| |||  
Db 478 AYIGFVHMVTPPPG-----VATGNWCGGAFG 504  
  
RESULT 11  
O33907 PRELIMINARY; PRT; 543 AA.  
AC O33907;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Shewanella sp. SCRC-2738.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=53560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SCRC-2738;  
RX MEDLINE=97419510; PubMed=9274025;  
RA Takeyama H., Takeda D., Yazawa K., Yamada A., Matsunaga T.;  
RT "Expression of the eicosapentaenoic acid synthesis gene cluster from  
RT Shewanella sp. in a transgenic marine cyanobacterium, Synechococcus  
RT sp.";  
RL Microbiology 143:0-0(0).  
DR EMBL; U73935; AAB81126.1; -.  
DR PIR; T30186; T30186.  
KW Hypothetical protein.  
SQ SEQUENCE 543 AA; 59378 MW; 7233F53635B794C7 CRC64;  
  
Query Match 39.8%; Score 60.5; DB 2; Length 543;  
Best Local Similarity 46.7%; Pred. No. 3.3;  
Matches 14; Conservative 4; Mismatches 9; Indels 3; Gaps 2;  
  
QY 2 YCGFLRPGVSSENLSAVATGNKXG--CGAFG 29  
||| : ||| ||| : ||| ||| |||  
Db 99 YAGAMANGISSEEL-VIALGQAGILCSFG 127  
  
RESULT 12  
O9SKB3 PRELIMINARY; PRT; 548 AA.  
AC O9SKB3; Q94ET7;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)



DE Putative poly(ADP-ribose) glycohydrolase.  
GN AT2G31870 OR TEJ.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,  
RA Barnstead M.E., Mason T.M., Bowman C.L., Romning C.M., Benito M.-I.,  
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,  
RA Fraser C.M., Venter J.C.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Panda S., Poirier G.G., Kay S.A.;  
RT "TEJ defines a role for poly-ADP-ribosylation in establishing period  
RT length of the Arabidopsis circadian oscillator";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006533; AAD32285.2; -.  
DR EMBL; AF394690; AAK72256.1; -.  
DR PIR; B84726; B84726.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 548 AA; 62169 MW; FLA79FDA157C3329 CRC64;

Query Match 39.5%; Score 60; DB 10; Length 548;  
Best Local Similarity 51.9%; Pred. No. 4;  
Matches 14; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY 5 FLRPGVSSNL--SAVATGNXGCGAFG 29  
| | | | |  
| | | | |  
Db 432 FYVEGVNEDHEDDGVATGNWCGVFG 458

RESULT 13  
Q8VYAI  
ID Q8VYAI PRELIMINARY; PRT; 522 AA.  
AC Q8VYAI;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Putative poly(ADP-ribose) glycohydrolase.  
GN AT2G31870.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY072330; AAL61937.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;

Query Match 37.5%; Score 57; DB 10; Length 522;  
Best Local Similarity 64.7%; Pred. No. 10;  
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 13 ENLSAVATGNXGCGAFG 29  
| | | | |  
| | | | |  
Db 424 EKKIGVATGNWCGVFG 440  
  
RESULT 14  
Q8DF10  
ID Q8DF10 PRELIMINARY; PRT; 513 AA.  
AC Q8DF10;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Na+/alanine symporter.  
GN VV10415.  
OS Vibrio vulnificus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=672;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CMCP6;  
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.;  
RT "Complete genome sequence of Vibrio vulnificus CMCP6";  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016798; AAO08938.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005283; F:sodium:amino acid transporter activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR GO; GO:0006814; P:sodium ion transport; IEA.  
DR InterPro; IPR002293; AA/rel\_permease1.  
DR InterPro; IPR001463; Na/Ala\_symp1.  
DR Pfam; PF01235; Na\_Ala\_symp; 1.  
DR PRINTS; PR00175; NAALASMPORT.  
DR TIGRFAMS; TIGR00835; agcs; 1.  
DR PROSITE; PS00873; NA\_ALANINE\_SYMP; 1.  
KW Complete proteome.  
SQ SEQUENCE 513 AA; 55137 MW; 545B3948378CF10D CRC64;

Query Match 36.5%; Score 55.5; DB 16; Length 513;  
Best Local Similarity 48.1%; Pred. No. 17;  
Matches 13; Conservative 2; Mismatches 7; Indels 5; Gaps 1;

QY 3 CGFLRPGVSSNL SAVATGNXGCGAFG 29  
| | | | |  
| | | | |  
Db 152 CGFLRPGVQNSI-----GNAVEAAFG 173

RESULT 15  
Q9VIS9  
ID Q9VIS9 PRELIMINARY; PRT; 578 AA.  
AC Q9VIS9;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG10366 protein.  
GN CG10366.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003663; AAF53836.1; -.  
DR FlyBase; FBgn0032814; CG10366.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf-C2H2; 8.  
DR SMART; SM00355; Znf\_C2H2; 8.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 8.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 578 AA; 65401 MW; 53F060B21E19010C CRC64;

Query Match 36.2%; Score 55; DB 5; Length 578;  
Best Local Similarity 57.9%; Pred. No. 22;  
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 11 SSENLSAVATGNXGCGAFG 29  
Db : ||| ||| ||| |||  
501 NKDNLSPVATKQSGCFAFG 519

Search completed: May 26, 2004, 18:46:10  
Job time : 4.31439 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 1.26028 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 53    | 100.0       | 11     | 5     | Aae25636 Bovine po |
| 2          | 53    | 100.0       | 11     | 5     | Aau76027 Bovine po |
| 3          | 53    | 100.0       | 11     | 5     | Aau76018 Bovine po |
| 4          | 53    | 100.0       | 11     | 6     | Abg72285 Oligopept |
| 5          | 53    | 100.0       | 968    | 5     | Aae25631 Murine po |
| 6          | 53    | 100.0       | 968    | 5     | Aau76022 Mouse pol |
| 7          | 53    | 100.0       | 968    | 5     | Aau76013 Mouse pol |
| 8          | 53    | 100.0       | 968    | 6     | Abg72280 Mmurine p |
| 9          | 53    | 100.0       | 977    | 5     | Aae25629 Bovine po |
| 10         | 53    | 100.0       | 977    | 5     | Aau76020 Bovine po |
| 11         | 53    | 100.0       | 977    | 5     | Aau75799 Bovine po |
| 12         | 53    | 100.0       | 977    | 6     | Abg72278 Bovine po |
| 13         | 52    | 98.1        | 976    | 5     | Aae25630 Human pol |
| 14         | 52    | 98.1        | 976    | 5     | Aau76021 Human pol |
| 15         | 52    | 98.1        | 976    | 5     | Aau76012 Human pol |
| 16         | 52    | 98.1        | 976    | 6     | Abg72279 Human pol |
| 17         | 47    | 88.7        | 10     | 5     | Aae25650 Bovine po |
| 18         | 41    | 77.4        | 249    | 6     | Abu26785 Protein e |
| 19         | 41    | 77.4        | 252    | 2     | Aay35276 Chlamydia |
| 20         | 37    | 69.8        | 103    | 3     | Aab40810 Human ORF |
| 21         | 37    | 69.8        | 103    | 5     | Abp02512 Human ORF |
| 22         | 37    | 69.8        | 1323   | 2     | Aar55248 N-methyl- |
| 23         | 37    | 69.8        | 1323   | 7     | Add48372 Rat Prote |
| 24         | 37    | 69.8        | 1324   | 7     | Add48915 Rat Prote |
| 25         | 37    | 69.8        | 1324   | 7     | Add47929 Rat Prote |

|    |    |      |      |   |          |                    |
|----|----|------|------|---|----------|--------------------|
| 26 | 37 | 69.8 | 1336 | 2 | AAR66041 | Aar66041 Human N-m |
| 27 | 37 | 69.8 | 1336 | 2 | AAW87511 | Aaw87511 Human N-m |
| 28 | 37 | 69.8 | 1336 | 3 | AAY56135 | Aay56135 Human NMD |
| 29 | 37 | 69.8 | 1336 | 3 | AAB26237 | Aab26237 Human N-m |
| 30 | 37 | 69.8 | 1336 | 5 | ABB56530 | Abb56530 Human NMD |
| 31 | 37 | 69.8 | 1336 | 5 | AAO18085 | Aao18085 Human NMD |
| 32 | 37 | 69.8 | 1336 | 6 | ABU65900 | Abu65900 Human N-m |
| 33 | 37 | 69.8 | 1336 | 6 | ABU61447 | Abu61447 Human N-m |
| 34 | 37 | 69.8 | 1336 | 6 | ABU57682 | Abu57682 Human N-m |
| 35 | 37 | 69.8 | 1336 | 6 | ABO43092 | AbO43092 Human NMD |
| 36 | 37 | 69.8 | 1336 | 7 | AAE39258 | Aae39258 Human NMD |
| 37 | 37 | 69.8 | 1336 | 7 | ADD48374 | Add48374 Human Pro |
| 38 | 36 | 67.9 | 123  | 6 | ADA34638 | Ada34638 Acinetoba |
| 39 | 36 | 67.9 | 519  | 2 | AAW70285 | Aaw70285 Oryza sat |
| 40 | 35 | 66.0 | 256  | 6 | ABM71099 | Abm71099 Staphyloc |
| 41 | 35 | 66.0 | 358  | 2 | AAW38601 | Aaw38601 S. pneumo |
| 42 | 35 | 66.0 | 449  | 6 | ABU00841 | Abu00841 S. pneumo |
| 43 | 35 | 66.0 | 449  | 6 | ABU45900 | Abu45900 Protein e |
| 44 | 35 | 66.0 | 527  | 6 | ABR53394 | Abr53394 Protein s |
| 45 | 35 | 66.0 | 1043 | 4 | ABB60095 | Abb60095 Drosophil |

ALIGNMENTS

RESULT 1  
AAE25636  
ID AAE25636 standard; peptide; 11 AA.  
XX  
AC AAE25636;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #3.  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 23  
FT /label= Unknown  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
XX  
DR New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
XX ribose glycohydrolase involved in cellular response to DNA damage,  
XX inhibition of which is useful for treating neoplastic disorders and  
XX neurodegenerative diseases.  
XX  
PS Claim 10; Col 81-82; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The

CC invention is useful as probes and primer molecules that can be used in  
CC hybridization assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is bPARG oligopeptide  
XX  
SQ Sequence 11 AA;  
Query Match 100.0%; Score 53; DB 5; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.00072;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FLINPELIVSR 11  
Db 1 FLINPELIVSR 11  
RESULT 2  
AAU76027  
ID AAU76027 standard; peptide; 11 AA.  
XX  
AC AAU76027;  
DT 08-MAY-2002 (first entry)  
XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 61.  
XX  
KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease; oligopeptide 61.  
XX  
OS Bos taurus.  
XX US6337202-B1.  
PN 08-JAN-2002.  
PD 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-163240/21.  
XX  
DR Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
PS Claim 4; Col 25; 81pp; English.  
XX  
CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.

CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents bovine PARG  
CC oligonucleotide 61. This peptide is one of several PARG oligopeptides  
CC (AAU76025-AAU76028) of the invention  
XX  
SQ Sequence 11 AA;  
Query Match 100.0%; Score 53; DB 5; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.00072;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FLINPELIVSR 11  
Db 1 FLINPELIVSR 11  
RESULT 3  
AAU76018  
ID AAU76018 standard; peptide; 11 AA.  
XX  
AC AAU76018;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 61.  
XX  
KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity; oligopeptide 61.  
XX  
OS Bos taurus.  
XX US6333148-B1.  
PN 25-DEC-2001.  
PD 30-APR-1999; 99US-00302812.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-153820/20.  
XX  
DR Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
PS Claim 4; Col 25; 80pp; English.  
XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents bovine PARG oligopeptide 61. This peptide is one of several  
CC PARG oligopeptides (AAU76016-AAU76019) of the invention  
XX  
SQ Sequence 11 AA;  
Query Match 100.0%; Score 53; DB 5; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.00072;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 FLINPELIVSR 11  
| | | | | | | |  
Db 1 FLINPELIVSR 11

RESULT 4  
ABG72285  
ID ABG72285 standard; peptide; 11 AA.  
XX  
AC ABG72285;  
XX  
DT 13-MAR-2003 (first entry)  
XX  
DE Oligopeptide #3 derived from bovine PARG enzyme.  
XX  
KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; neotropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective.  
XX  
OS Bos taurus.  
XX  
PN US2002132328-A1.  
XX  
PD 19-SEP-2002.  
XX  
PF 09-OCT-2001; 2001US-00973451.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2003-155895/15.  
DR  
XX  
XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
PS Example 2; Page 14; 86pp; English.  
XX  
CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. ABG72283-  
CC ABG72286 represent oligopeptides derived from bovine PARG enzyme. The  
CC oligopeptides are used to construct degenerate PCR primers for the  
CC isolation of cDNA encoding bovine PARG  
XX  
SQ Sequence 11 AA;  
Query Match 100.0%; Score 53; DB 6; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.00072; Indels 0; Gaps 0;  
Matches 11; Conservative 0; Mismatches 0;  
QY 1 FLINPELIVSR 11  
| | | | | | | |  
Db 1 FLINPELIVSR 11

RESULT 5  
AAE25631  
ID AAE25631 standard; protein; 968 AA.  
XX  
AC AAE25631;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
XX  
KW Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Mus musculus.  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-535641/57.  
DR N-PSDB; AAD42083.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Claim 3; Col 63-68; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is murine PARG  
XX  
SQ Sequence 968 AA;  
Query Match 100.0%; Score 53; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;  
Matches 11; Conservative 0; Mismatches 0;  
QY 1 FLINPELIVSR 11  
| | | | | | | |  
Db 751 FLINPELIVSR 761

RESULT 6  
AAU76022  
ID AAU76022 standard; protein; 968 AA.  
XX  
AC AAU76022;  
XX  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurofugal disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
OS Mus musculus.  
XX  
PN US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-163240/21.  
DR N-PSDB; ABK14933.  
XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
PS Claim 2; Col 63-70; 81pp; English.  
XX  
CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders, including Alzheimer's, Huntington's and Parkinson's diseases,  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the mouse PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX  
SQ Sequence 968 AA;  
  
Query Match 100.0%; Score 53; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FLINPELIVSR 11  
Db 751 FLINPELIVSR 761  
  
RESULT 7  
AAU76013  
ID AAU76013 standard; protein; 968 AA.

XX  
AC AAU76013;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX  
OS Mus musculus.  
XX  
PN US6333148-B1.  
XX  
PD 25-DEC-2001.  
XX  
PF 30-APR-1999; 99US-00302812.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-153820/20.  
DR N-PSDB; ABK14495.  
XX  
PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
PS Claim 3; Col 63-68; 80pp; English.  
XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the mouse PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX  
SQ Sequence 968 AA;  
  
Query Match 100.0%; Score 53; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FLINPELIVSR 11  
Db 751 FLINPELIVSR 761  
  
RESULT 8  
ABG72280  
ID ABG72280 standard; protein; 968 AA.  
XX  
AC ABG72280;  
XX  
DT 13-MAR-2003 (first entry)  
XX  
DE Murine poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX  
KW Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;

KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
KW cytosstatic; neuroprotective; nootropic; antiparkinsonian; cardiant;  
KW vasotropic; anticonvulsant; cerebroprotective; enzyme.  
XX  
OS Mus musculus.  
XX  
XX  
PN US2002132328-A1.  
XX  
PD 19-SEP-2002.  
XX  
XX 09-OCT-2001; 2001US-00973451.  
PF  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2003-155895/15.  
DR N-PSDB; ABX14479.  
XX  
XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
XX Claim 28; Fig 16; 86pp; English.  
XX  
XX The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents murine PARG enzyme  
XX  
SQ Sequence 968 AA;  
  
Query Match 100.0%; Score 53; DB 6; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FLINPELIVSR 11  
DB 751 FLINPELIVSR 761  
  
RESULT 9  
AAE25629  
ID AAE25629 standard; protein; 977 AA.  
XX  
AC AAE25629;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
XX

KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytosstatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
XX 23-FEB-2000; 2000US-00511507.  
PF  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
PA  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-535641/57.  
DR N-PSDB; AAD42081.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
XX Claim 3; Col 47-45; 77pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX  
SQ Sequence 977 AA;  
  
Query Match 100.0%; Score 53; DB 5; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FLINPELIVSR 11  
DB 760 FLINPELIVSR 770  
  
RESULT 10  
AAU76020  
ID AAU76020 standard; protein; 977 AA.  
XX  
AC AAU76020;  
XX  
XX 08-MAY-2002 (first entry)  
DT  
XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.

XX OS Bos taurus.  
XX FH Key  
FT FT Location/Qualifiers  
FT FT 601..617  
FT FT /note= "Represents PARG oligopeptide #1"  
FT FT Region  
FT FT 761..770  
FT FT /note= "Represents PARG oligopeptide #2"  
FT FT Region  
FT FT 771..801  
FT FT /note= "Represents PARG oligopeptide #3"  
FT FT Region  
FT FT 849..880  
FT FT /note= "Represents PARG oligopeptide #4"  
XX XX  
PN US6337202-B1.  
XX XX  
PD 08-JAN-2002.  
XX XX  
PF 23-FEB-2000; 2000US-00511477.  
XX XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-163240/21.  
DR N-PSDB; ABK14931.  
XX XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX XX  
PS Claim 2; Col 47-52; 81pp; English.  
XX XX  
CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders, including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the bovine PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX XX  
SQ Sequence 977 AA;

Query Match 100.0%; Score 53; DB 5; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
| | | | | | | | | |  
Db 760 FLINPELIVSR 770

RESULT 11  
AAU75799  
ID AAU75799 standard; protein; 977 AA.  
XX XX  
AC AAU75799;  
XX XX  
DT 08-MAY-2002 (first entry)  
XX XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX XX  
KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;

KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX OS Bos taurus.  
XX FH Key  
FT FT Location/Qualifiers  
FT FT 601..617  
FT FT /note= "Represents PARG oligopeptide #1"  
FT FT Region  
FT FT 761..770  
FT FT /note= "Represents PARG oligopeptide #2"  
FT FT Region  
FT FT 771..801  
FT FT /note= "Represents PARG oligopeptide #3"  
FT FT Region  
FT FT 849..880  
FT FT /note= "Represents PARG oligopeptide #4"  
XX XX  
PN US6333148-B1.  
XX XX  
PD 25-DEC-2001.  
XX XX  
PF 30-APR-1999; 99US-00302812.  
XX XX  
PR 01-MAY-1998; 98US-0083768P.  
XX XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-153820/20.  
DR N-PSDB; ABK14493.  
XX XX  
PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX XX  
PS Claim 3; Col 45-52; 80pp; English.

XX XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the bovine PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX XX  
SQ Sequence 977 AA;

Query Match 100.0%; Score 53; DB 5; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
| | | | | | | | | |  
Db 760 FLINPELIVSR 770

RESULT 12  
ABG72278  
ID ABG72278 standard; protein; 977 AA.  
XX XX  
AC ABG72278;  
XX XX  
DT 13-MAR-2003 (first entry)  
XX XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX XX



KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; neoplastic; antiparkinsonian; cardiant; vasotrophic;  
KW anticonvulsant; cerebroprotective; enzyme.  
XX  
OS Bos taurus.  
XX  
PN US2002132328-A1.  
XX  
PD 19-SEP-2002.  
XX  
PF 09-OCT-2001; 2001US-00973451.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2003-155895/15.  
DR N-PSDB; ABX14477.  
XX  
XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
XX Claim 28; Fig 16; 86pp; English.  
PS  
XX The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents bovine PARG enzyme  
XX  
SQ Sequence 977 AA;

Query Match 100.0%; Score 53; DB 6; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||||  
Db 760 FLINPELIVSR 770

RESULT 13  
AAE25630  
ID AAE25630 standard; protein; 976 AA.

XX AAE25630;

XX 04-NOV-2002 (first entry)

XX

DE Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
XX  
KW Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotrophic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
XX Homo sapiens.  
OS  
XX US6395543-B1.  
PN  
XX 28-MAY-2002.  
PD  
XX 23-FEB-2000; 2000US-00511507.  
PF  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
PA  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI  
XX WPI; 2002-535641/57.  
DR N-PSDB; AAD42082.  
XX

New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

Claim 3; Col 55-60; 77pp; English.

The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The antisense therapeutic are useful to treat neoplastic disorders and conditions caused by genotoxic oxidative stress e.g., cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. The invention is useful in gene therapy and antisense therapy.

The present sequence is human PARG

Sequence 976 AA;  
Query Match 98.1%; Score 52; DB 5; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.23;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||||  
Db 759 FLINPELIVSR 769

RESULT 14  
AAU76021  
ID AAU76021 standard; protein; 976 AA.

XX AAU76021;

XX 08-MAY-2002 (first entry)

XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW

KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.

XX Homo sapiens.

PN US6337202-B1.

XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-00511477.

XX 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

PA Jacobson MK, Jacobson EL, Ame J, Lin W;

PI WPI; 2002-163240/21.

XX N-PSDB; ABK14932.

XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 2; Col 55-60; 81pp; English.

XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the human PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention

XX Sequence 976 AA;

Query Match 98.1%; Score 52; DB 5; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.23;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11

Db 759 FLINPELLIISR 769

RESULT 15

AAU76012  
ID AAU76012 standard; protein; 976 AA.

XX AC AAU76012;

XX 08-MAY-2002 (first entry)

XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.

XX OS Homo sapiens.

XX PN US6333148-B1.

XX

PD 25-DEC-2001.

XX 30-APR-1999; 99US-00302812.

PF 01-MAY-1998; 98US-0083768P.

PR (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

PI WPI; 2002-153820/20.

XX N-PSDB; ABK14494.

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.

XX Claim 3; Col 55-60; 80pp; English.

XX The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage,  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the human PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention

XX Sequence 976 AA;

Query Match 98.1%; Score 52; DB 5; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.23;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11

Db 759 FLINPELLIISR 769

Search completed: May 26, 2004, 18:40:12  
Job time : 4.26028 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 77.2527 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-6  
Perfect score: 5164  
Sequence: 1 MSAGPGWEPTKARWGAAAGT.....YHAVESSAETDMPGQKAGT 968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriaph:\*
  - 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1          | 5164   | 100.0       | 968    | 11    | O88622      |
| 2          | 5098.5 | 98.7        | 961    | 11    | Q80YQ6      |
| 3          | 4869.5 | 94.3        | 920    | 11    | Q8CB72      |
| 4          | 4749   | 92.0        | 972    | 11    | Q9QYM2      |
| 5          | 4456   | 86.3        | 976    | 4     | Q86W56      |
| 6          | 4451   | 86.2        | 976    | 4     | Q7Z742      |
| 7          | 4395   | 85.1        | 976    | 4     | Q9Y4W7      |
| 8          | 4332.5 | 83.9        | 977    | 6     | O02776      |
| 9          | 1074.5 | 20.8        | 768    | 5     | O46043      |
| 10         | 1069.5 | 20.7        | 723    | 5     | Q960N8      |
| 11         | 568    | 11.0        | 548    | 10    | Q9SKB3      |
| 12         | 525    | 10.2        | 522    | 10    | Q8VYAL      |
| 13         | 458    | 8.9         | 485    | 5     | Q9N5L4      |
| 14         | 451.   | 8.7         | 781    | 5     | Q867X0      |
| 15         | 450.5  | 8.7         | 764    | 5     | Q19637      |
| 16         | 328.5  | 6.4         | 368    | 5     | Q86GI4      |

|    |       |     |      |    |        |                     |
|----|-------|-----|------|----|--------|---------------------|
| 17 | 328   | 6.4 | 364  | 10 | Q9SKB4 | Q9skb4 arabidopsis  |
| 18 | 183   | 3.5 | 1194 | 11 | Q8C0H4 | Q8c0h4 mus musculus |
| 19 | 183   | 3.5 | 1194 | 11 | Q8BMO0 | Q8bmo0 mus musculus |
| 20 | 179.5 | 3.5 | 1193 | 11 | Q8BTQ4 | Q8btq4 mus musculus |
| 21 | 178.5 | 3.5 | 931  | 5  | Q9VF92 | Q9vtf92 drosophila  |
| 22 | 177   | 3.4 | 930  | 5  | Q8IHH1 | Q8ihh1 drosophila   |
| 23 | 177   | 3.4 | 5322 | 5  | Q9VPL9 | Q9vpl9 drosophila   |
| 24 | 176.5 | 3.4 | 888  | 12 | Q8JKL5 | Q8jkl5 heliothis z  |
| 25 | 176   | 3.4 | 8081 | 5  | Q7Z120 | Q7z120 caenorhabdi  |
| 26 | 174.5 | 3.4 | 571  | 5  | Q8MTN7 | Q8mtn7 trichinella  |
| 27 | 172   | 3.3 | 508  | 12 | Q9E234 | Q9e234 helicoverpa  |
| 28 | 168   | 3.3 | 3166 | 5  | Q9W3Z0 | Q9w3z0 drosophila   |
| 29 | 167   | 3.2 | 510  | 12 | Q99GU9 | Q99gu9 helicoverpa  |
| 30 | 167   | 3.2 | 1169 | 16 | Q7ZAJ1 | Q7zaj1 staphylococ  |
| 31 | 165   | 3.2 | 498  | 5  | Q8MTN8 | Q8mtn8 trichinella  |
| 32 | 165   | 3.2 | 1192 | 5  | Q96127 | Q96127 plasmodium   |
| 33 | 164.5 | 3.2 | 1537 | 4  | Q96NW7 | Q96nw7 homo sapien  |
| 34 | 163   | 3.2 | 532  | 10 | Q9T069 | Q9t069 arabidopsis  |
| 35 | 162   | 3.1 | 1401 | 2  | Q8RJR4 | Q8rjr4 staphylococ  |
| 36 | 162   | 3.1 | 4969 | 11 | Q8CF91 | Q8cf91 mus musculus |
| 37 | 162   | 3.1 | 5165 | 11 | Q8CF92 | Q8cf92 mus musculus |
| 38 | 161.5 | 3.1 | 678  | 13 | Q803F8 | Q803f8 brachydanio  |
| 39 | 161.5 | 3.1 | 2392 | 5  | Q810K1 | Q810k1 caenorhabdi  |
| 40 | 161.5 | 3.1 | 2407 | 5  | Q7YSH5 | Q7ysh5 caenorhabdi  |
| 41 | 160.5 | 3.1 | 1444 | 5  | Q9VTN2 | Q9vtn2 drosophila   |
| 42 | 160.5 | 3.1 | 1514 | 5  | Q8SY55 | Q8sy55 drosophila   |
| 43 | 160.5 | 3.1 | 1566 | 5  | Q8IAD6 | Q8iad6 plasmodium   |
| 44 | 159.5 | 3.1 | 805  | 4  | Q8TF01 | Q8tf01 homo sapien  |
| 45 | 159.5 | 3.1 | 1115 | 5  | Q17863 | Q17863 caenorhabdi  |

ALIGNMENTS

RESULT 1

O88622 ID O88622 PRELIMINARY; PRT; 968 AA.  
AC O88622;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9380098; PubMed=10449915;  
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to human chromosome 10q11.23 and mouse chromosome 14B by in situ hybridization.";  
RT Cytogenet. Cell Genet. 85:269-270(1999).  
RL EMBL; AF079557; AAC28735.1; -.  
DR MGD; MGI:1347094; Parg.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;

Query Match 100.0%; Score 5164; DB 11; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |    |   |
|----|----|---|
| Qy | 1  | MSAGPGWEPTKARWGAAAGTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60 |
| Db | 1  | MSAGPGWEPTKARWGAAAGTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60 |
| Qy | 61 | RAGPHRGNATSFVFKQKTIITWMDTKGPKTAESKNNTRIDSMSSVQKDNFYPHKV 120 |

|          |   |   |              |
|----------|---|---|--------------|
| Db       | 61  | RAGPHRGNATSFVFKQKTTITWMDTKGPKTAESKESKENNNTRIDSMSSVQKDNFYPHKV  | 120          |
| QY       | 121   | EKLENVQNLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTLPKQLS   | 180          |
| Db       | 121   | EKLENVQNLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTLPKQLS   | 180          |
| QY       | 181   | NANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQARSNCCKSGSRQSVKD  | 240          |
| Db       | 181   | NANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQARSNCCKSGSRQSVKD  | 240          |
| QY       | 241   | CTGCQEQEVDVLPESPLSDVGAEDIGTGPKNDKLTGQESSLGDSPPFEKESPEPMDV     | 300          |
| Db       | 241   | CTGCQEQEVDVLPESPLSDVGAEDIGTGPKNDKLTGQESSLGDSPPFEKESPEPMDV     | 300          |
| QY       | 301   | DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTGSEVRLH   | 360          |
| Db       | 301   | DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTGSEVRLH   | 360          |
| QY       | 361   | FQFEGENNAGTSDLNAPSGNSSSLNVECRSSKQHGKRDSTKIDHFMRIKSEDRRKEQC    | 420          |
| Db       | 361   | FQFEGENNAGTSDLNAPSGNSSSLNVECRSSKQHGKRDSTKIDHFMRIKSEDRRKEQC    | 420          |
| QY       | 421   | EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPRLSPSASHTVTVRVDL  | 480          |
| Db       | 421   | EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPRLSPSASHTVTVRVDL  | 480          |
| QY       | 481   | LRAGEVPKPPFTHYKDLWKNHVKMPCSEQNLVPEDENGERTAGSRWELIQTALLNKFT    | 540          |
| Db       | 481   | LRAGEVPKPPFTHYKDLWKNHVKMPCSEQNLVPEDENGERTAGSRWELIQTALLNKFT    | 540          |
| QY       | 541   | RPQNLKDAILKYNVAYSKKWDFALTALVDFWDKVLLEAEQAHLQYQILPDMVKIALCLPNI | 600          |
| Db       | 541   | RPQNLKDAILKYNVAYSKKWDFALTALVDFWDKVLLEAEQAHLQYQILPDMVKIALCLPNI | 600          |
| QY       | 601   | TPQIPLLKQKMNHVSVMQEQIASLLANAFCTFPRRNMKSEYSSYPDINFNRLFEGR      | 660          |
| Db       | 601   | TPQIPLLKQKMNHVSVMQEQIASLLANAFCTFPRRNMKSEYSSYPDINFNRLFEGR      | 660          |
| QY       | 661   | SSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDFFPWERCEKPLTRLHVTYEGTIEG    | 720          |
| Db       | 661   | SSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDFFPWERCEKPLTRLHVTYEGTIEG    | 720          |
| QY       | 721   | NGRGLQVDFANRFVGGVGTAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ    | 780          |
| Db       | 721   | NGRGLQVDFANRFVGGVGTAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ    | 780          |
| QY       | 781   | YSEYTGAEYRWARSHEDGSEKDDWQRRCCTEIVADALHFRYRLDQFVPEKVRRELNKA    | 840          |
| Db       | 781   | YSEYTGAEYRWARSHEDGSEKDDWQRRCCTEIVADALHFRYRLDQFVPEKVRRELNKA    | 840          |
| QY       | 841   | YCGFLRPGVPSENLSAVATGNWGCAGFGGDARLKALIQILAAAAAERDVYFTFGDSELM   | 900          |
| Db       | 841   | YCGFLRPGVPSENLSAVATGNWGCAGFGGDARLKALIQILAAAAAERDVYFTFGDSELM   | 900          |
| QY       | 901   | RDIYSMHTFLTERKLDVGKVKYLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD  | 960          |
| Db       | 901   | RDIYSMHTFLTERKLDVGKVKYLLRLRYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD  | 960          |
| QY       | 961   | MPGQKAGT 968  |              |
| Db       | 961   | MPGQKAGT 968  |              |
| RESULT 2 |   |   |              |
| Q80YQ6   |   |   |              |
| ID       | Q80YQ6  | PRELIMINARY;  | PRT; 961 AA. |
| AC       | Q80YQ6;   |   |              |
| DT       | 01-JUN-2003   | (TrEMBLrel. 24, Created)                                      |              |
| DT       | 01-JUN-2003   | (TrEMBLrel. 24, Last sequence update)                         |              |
| DT       | 01-OCT-2003   | (TrEMBLrel. 25, Last annotation update)                       |              |
| DE       | Poly  | (ADP-ribose) glycohydrolase.                                  |              |
| OS       | Mus musculus  | (Mouse).  |              |
| OC       | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |   |              |

|  |  |
|--|--|
| OC   | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.     |
| OX   | NCBI_TaxID=10090;  |
| RN   | [1]  |
| RP   | SEQUENCE FROM N.A.   |
| RC   | STRAIN=C57BL/6; TISSUE=Brain;  |
| RA   | Strausberg R.;   |
| RL   | Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.                |
| DR   | EMBL; BC050892; AAH50892.1; -.   |
| DR   | GO; GO:0016787; F:hydrolase activity; IEA.                             |
| DR   | InterPro; IPR007724; PARG.   |
| DR   | Pfam; PF05028; PARG; 1.  |
| KW   | Hydrolase.   |
| SQ   | SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;                    |
| Query Match 98.7%; Score 5098.5; DB 11; Length 961;          |  |
| Best Local Similarity 99.6%; Pred. No. 0;                    |  |
| Matches 957; Conservative 1; Mismatches 2; Indels 1; Gaps 1; |  |
| QY   | 1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSPACVS 59       |
| Db   | 1 MSAGPGWEPCTKRPRWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSPACVS 60       |
| QY   | 60 GRAGPHRGNATSFVKQKTTITWMDTKGPKTAESKESKENNNTRIDSMSSVQKDNFYPHK 119     |
| Db   | 61 GRAGPHRGNATSFVKQKTTITWMDTKGPKTAESKESKENNNTRIDSMSSVQKDNFYPHK 120     |
| QY   | 120 VEKLENVQNLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTLPKQL 179    |
| Db   | 121 VEKLENVQNLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTLPKQL 180    |
| QY   | 180 SNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQARSNCCKSGSRQSVK 239   |
| Db   | 181 SNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQARSNCCKSGSRQSVK 240   |
| QY   | 240 DCTGCQEQEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESPEPMD 299     |
| Db   | 241 DCTGCQEQEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESPEPMD 300     |
| QY   | 300 VDNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTGSEVRL 359    |
| Db   | 301 VDNSKNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTGSEVRL 360    |
| QY   | 360 HFQFEGENNAGTSDLNAPSGNSSSLNVECRSSKHQKRDSTKIDHFMRIKSEDRRKEQ 419      |
| Db   | 361 HFQFEGENNAGTSDLNAPSGNSSSLNVECRSSKHQKRDSTKIDHFMRIKSEDRRKEQ 420      |
| QY   | 420 CEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPRLSPSASHTVTVRVD 479   |
| Db   | 421 CEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPRLSPSASHTVTVRVD 480   |
| QY   | 480 LLRAGEVPKPPFTHYKDLWKNHVKMPCSEQNLVPVEDENGERTAGSRWELIQTALLNKF 539    |
| Db   | 481 LLRAGEVPKPPFTHYKDLWKNHVKMPCSEQNLVPVEDENGERTAGSRWELIQTALLNKF 540    |
| QY   | 540 TRPQNLKDAILKYNVAYSKKWDFALTALVDFWDKVLLEAEQAHLQYQILPDMVKIALCLPNI 599 |
| Db   | 541 TRPQNLKDAILKYNVAYSKKWDFALTALVDFWDKVLLEAEQAHLQYQILPDMVKIALCLPNI 600 |
| QY   | 600 CTQPIPLKQKMNHVSVMQEQIASLLANAFCTFPRRNMKSEYSSYPDINFNRLFEG 659        |
| Db   | 601 CTQPIPLKQKMNHVSVMQEQIASLLANAFCTFPRRNMKSEYSSYPDINFNRLFEG 660        |
| QY   | 660 RSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDFFPWERCEKPLTRLHVTYEGTIE 719     |
| Db   | 661 RSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDFFPWERCEKPLTRLHVTYEGTIE 720     |
| QY   | 720 NGRGLQVDFANRFVGGVGTAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTE 779      |
| Db   | 721 NGRGLQVDFANRFVGGVGTAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTE 780      |
| QY   | 780 QYSEYTGAEYRWARSHEDGSEKDDWQRRCCTEIVADALHFRYRLDQFVPEKVRRELNK 839     |
| Db   | 781 QYSEYTGAEYRWARSHEDGSEKDDWQRRCCTEIVADALHFRYRLDQFVPEKVRRELNK 840     |



QY 840 AYCGLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSEL 899  
Db 841 AYCGLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSEL 900  
QY 900 MRDIYSMHTFLTERKLDVGKVKYKLLRLRYNNEECNCSTPGPDIKLYPFYHAVESSAETT 959  
Db 901 MRDIYSMHTFLTERKLDVGKVKYKLLRLRYNNEECNCSTPGPDIKLYPFYHAVESSAETT 960  
QY 960 D 960  
Db 961 D 961  
RESULT 3  
Q8CB72 PRELIMINARY; PRT; 920 AA.  
AC Q8CB72  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Poly.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;  
Query Match 94.3%; Score 4869.5; DB 11; Length 920;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 915; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
QY 1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59  
Db 1 MSAGPGWEPCTKPRRWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 60  
QY 60 GRAGPHRGNATSFVFKQKTIITWMDTKGPKTAESKESKENNNTRIDSMSSVQKDNFYPHK 119  
Db 61 GRAGPHRGNATSFVFKQKTIITWMDTKGPKTAESKESKENNNTRIDSMSSVQKDNFYPHK 120  
QY 120 VEKLENVQNLNLDKSPTEKSSQYLNQOQTASVCKWQNEGKHAEQLLASEPPAGTLPKQL 179  
Db 121 VEKLENVQNLNLDKSPTEKSSQYLNQOQTASVCKWQNEGKHAEQLLASEPPAGTLPKQL 180  
QY 180 SNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQARSNCCKSGSRQSVK 239  
Db 181 SNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQARSNCCKSGSRQSVK 240  
QY 240 DCTGCGQEEVDVLPESPLSDVGAEDICTGPKNDNKLTGQESSLGDSPPFEKESEPESEMD 299  
Db 241 DCTGCGQEEVDVLPESPLSDVGAEDICTGPKNDNKLTGQESSLGDSPPFEKESEPESEMD 300  
QY 300 VDNSRNSQDSEADEETSPVDEQDDRSSQTANKLSSQAREADGDLRKRYLTGKSEVRL 359  
Db 301 VDNSKNSQDSEADEETSPVDEQDDRSSQTANKLSSQAREADGDLRKRYLTGKSEVRL 360  
QY 360 HFQFEGENNAGTSDLNAPSGNSSSLNVECRSSQKHGKRDSKITDHFMRISKSEDRRKEQ 419  
Db 361 HFQFEGENNAGTSDLNAPSGNSSSLNVECRSSQKHGKRDSKITDHFMRISKSEDRRKEQ 420

QY 420 CEVRHQRTERRKIPKYPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVD 479  
Db 421 CEVRHQRTERRKIPKYPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVD 480  
QY 480 LLRAGEVPKPPPTHYKDLNDKHKVMPCEQNLYPVEDENGERTAGSRWELIQTALLNKF 539  
Db 481 LLRAGEVPKPPPTHYKDLNDKHKVMPCEQNLYPVEDENGERTAGSRWELIQTALLNKF 540  
QY 540 TRPQNLKDALIKYNVAYSKKWDFALTALVDFWDKVLAEAAQHLYSILPDMVKIALCLPNI 599  
Db 541 TRPQNLKDALIKYNVAYSKKWDFALTALVDFWDKVLAEAAQHLYSILPDMVKIALCLPNI 600  
QY 600 CTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAMKSEYSSYPDINFENRLFEG 659  
Db 601 CTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAMKSEYSSYPDINFENRLFEG 660  
QY 660 RSSRKPEKLTLCYFRRVTEKKPTGLVTFTTQSLSEDFEWEWERCEKPLTRLHVTYEGTIE 719  
Db 661 RSSRKPEKLTLCYFRRVTEKKPTGLVTFTTQSLSEDFEWEWERCEKPLTRLHVTYEGTIE 720  
QY 720 GNGRGMLOVDFANRFVGGVGTAGLVQEEIRFELINPELIVSRLFTFVLDHNECLLIITGTE 779  
Db 721 GNGRGMLOVDFANRFVGGVGTAGLVQEEIRFELINPELIVSRLFTFVLDHNECLLIITGTE 780  
QY 780 QYSEYTGAYETRWARSHEDGSEKDDWQRRTTEIVAIDALHFRRYLDQFVPEKVRRELNK 839  
Db 781 QYSEYTGAYETRWARSHEDGSEKDDWQRRTTEIVAIDALHFRRYLDQFVPEKVRRELNK 840  
QY 840 AYCGLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSEL 899  
Db 841 AYCGLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSEL 900  
QY 900 MRDIYSMHTFLTERKLDVGK 919  
Db 901 MRDIYSMHTFLTERKLDVGK 920  
RESULT 4  
Q9QYM2 PRELIMINARY; PRT; 972 AA.  
ID Q9QYM2  
AC Q9QYM2;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN PARG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BUF; TISSUE=Colon;  
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,  
RA Aoki Y., Nakagawa H., Sugimura T.;  
RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase  
RT (Parg).";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB019366; BAA87901.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;  
Query Match 92.0%; Score 4749; DB 11; Length 972;  
Best Local Similarity 92.0%; Pred. No. 0;  
Matches 894; Conservative 26; Mismatches 48; Indels 4; Gaps 2;  
QY 1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59  
Db 1 MSAGPGWEPCTKPRRWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 60

QY 60 GRAGPHRGNATSFVFKQKTTITWMDTKGPKTAESSEKENNNTRIDSMSSVQKDNFYPHK 119  
Db 61 GRAGPHRGVTSFVFKQKPIITWMDTKGPKTAESSEKENNNTRIDPMSSVQKDNFYPHK 120  
QY 120 VEKLENVQNLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGKHAELQASEPPAGTLPKQL 179  
Db 121 VEKLENVQNLNLDKSPTEKSTPYLNQOQTAGVCKWHSAGERAELQASEPSAVTQAPKQL 180  
QY 180 SNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVK 239  
Db 181 SNANIDQSPPTDGHSDTDHEEDRDNQOFLTPVKLANAKQTVGDGQARSNCKCSASCQCGQ 240  
QY 240 DCTGCQEEVDVLPESPLSDVGAEDIGTGPKNDNKLGTQESSLGSPPEKESEPESPMD 299  
Db 241 DCAGCQREADVLPESPLSDVGAEDIGTGSKNDNKLGTQESGLGSPPEKESEPESPMD 300  
QY 300 VDNRSNCQDSEADEETSVPFDE---QDDRSSQTANKLSSCQAREADGDLRKRYLTGSE 356  
Db 301 VDNKSTSCQDSEADEEASVPFDEQDDQDDRSSQTANKLSSRQAREVDGDLRKRYLTGSE 360  
QY 357 VRLHFQFEGENNAGTSDLNAPKPSGNSSSLNVECRSSKQHGKRDSKI THDFMRISKSEDRR 416  
Db 361 IRLHFQFEGGSNAGTSDLNAPKPSGNSSSLNVDGRSSKQHGKRDSKI THDFVRIPKSEDKR 420  
QY 417 KEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPTEEMRMKPRCGIHLPSLRPSASHTVTV 476  
Db 421 KEQCEVRHQAERKIPKYVPPNLPDPDKKWLGTPIEMRMKPRCGVRLPLLRPSASHTVTV 480  
QY 477 RVDLLRAGEVPKPPFTHYKDLWKNKHVMPCSEQNLYPVEDENGERTAGSRWELIQTALL 536  
Db 481 RVDLLRAGEVPKPPFTHYKDLWKNKHVMPCSEQNLYPVEDENGERTAGSRWELIQTALL 540  
QY 537 NKFTRPQNLKDAILKYNVAYSKKWDFALTALVDFWVKLEEAQAHLQYQSILPDMVKIALCL 596  
Db 541 NKFTRPQNLKDAILKYNVAYSKKWDFALTALVDFWVKLEEAQAHLQYQSILPDMVKIALCL 600  
QY 597 PNICTQPIPLKQKMNHSVTMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRL 656  
Db 601 PNICTQPIPLKQKMNHSVTMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRL 660  
QY 657 FEGRSRKPEKLTLCFYFRRVTEKKPTGLVTFTRQSLDPPPEWERCEKPLTRLHVTEG 716  
Db 661 FEGRSRKPEKLTLCFYFRRVTEKKPTGLVTFTRQSLDPPPEWERCEKPLTRLHVTEG 720  
QY 717 TIEGNRGMQLQVDFANRFVGGVGTGAGLVQEEIRFLINPELIVSRLEFTEVLDHNECLIIT 776  
Db 721 TIEGNRGMQLQVDFANRFVGGVGTGAGLVQEEIRFLINPELIVSRLEFTEVLDHNECLIIT 780  
QY 777 GTEQYSEYTGAEYRWARSHEDGSEKDDWQRCTEIVADALHFRYLDQFVPEKVRRE 836  
Db 781 GTEQYSEYTGAEYRWARSHEDGSEKDDWQRCTEIVADALHFRYLDQFVPEKVRRE 840  
QY 837 LNKAYCGFLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGD 896  
Db 841 LNKAYCGFLRPGVPPNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGD 900  
QY 897 SELMRDIYSMTFLTERKLDVGKYKLLRLRYNEECRNCSFPGDICKLYPFIYHAEVSSA 956  
Db 901 SELMRDIYSMTFLTERKLVNKGKYLRLRLRYREECRDCSSPGDICKLYPFIYHAEVSSA 960  
QY 957 ETTDMPGQKAGT 968  
Db 961 ETSQDPGQRTGT 972

RESULT 5  
Q86W56  
ID Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
Query Match 86.3%; Score 4456; DB 4; Length 976;  
Best Local Similarity 86.6%; Pred. No. 0;  
Matches 846; Conservative 49; Mismatches 72; Indels 10; Gaps 6;  
QY 1 MSAGPGWEPTK-ARWGAAGTSAPTASDSRSPGRQRRVLDPKADPVQFRVPPSPACVS 59  
Db 1 MNAGPGCEPCTKRPRWGAA-TTSPAASDARSFSPRQRRVLDPKDAHVQFRVPPSPACVP 59  
QY 60 GRAGPHRGNATSFVFKQKTTITWMDTKGPKTAESE---SKENNTRIDSMSSVQKDNFY 116  
Db 60 GRAGQHRGSATSLVFKQKTTITWMDTKGPKTAESSESLDSKENNNTRIESMMSSVQKDNFY 119  
QY 117 PHKVEKLENVQNLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGKHAELQASEPPAGTLP 176  
Db 120 QHNVEKLENVQNLNLDKSPTEKSTQYLNQOQTAAAMCKWQNEKGKHAELQASEPPAGTLP 179  
QY 177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDQAR---SNCKCSG 233  
Db 180 EQFSNANIDRSPQNDHSDTDSEENRDNQOFLTPVKLANAKQTTTDEQAREAKSHQKCSK 239  
QY 234 SRQSVKDCGCOQEEVDVLPESPLSDVGAEDIGTGPKNDNKLGTQESSLGSPPEKESE 293  
Db 240 SCDPGEDECASCQOQDEIDVVPESPLSDVGSSEVDVGTGPKNDNKLTRQESCLGNSPPPEKESE 299  
QY 294 PESPMVDVNSRNSCQDSEADEETSVPFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352  
Db 300 PESPMVDVNSKNSCQDSEADEETSVPFDEQDEQDSSSTANKPSRFOARDADIEFRKRYST 359  
QY 353 KGSEVRLHFQFE-GENNAGTSDLNAPKPSGNSSSLNVECRSSKQHGKRDSKI THDFMRISK 411  
Db 360 KGSEVRLHFQFEGESRTGMNDLNKALPKNLPGNISLNVNVECRNSKQHGKRDSKI THDFMRILPK 419  
QY 412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPTEEMRMKPRCGIHLPSLRPSAS 471  
Db 420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPTEEMRMKPRCGIHLPSLRPSAN 479  
QY 472 HTVTVRVDLLRAGEVPKPPFTHYKDLWKNKHVMPCSEQNLYPVEDENGERTAGSRWELI 531  
Db 480 HTVTIRVDLLRAGEVPKPPFTHYKDLWKNKHVMPCSEQNLYPVEDENGERTAGSRWELI 539  
QY 532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFALTALVDFWVKLEEAQAHLQYQSILPDMVK 591  
Db 540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFALTALVDFWVKLEEAQAHLQYQSILPDMVK 599  
QY 592 IALCLPNICTQPIPLKQKMNHSVTMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 651  
Db 600 IALCLPNICTQPIPLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659  
QY 652 NFNRLEFEGRSRKPEKLTLCFYFRRVTEKKPTGLVTFTRQSLDPPPEWERCEKPLTRLH 711  
Db 660 NFNRLEFEGRSRKPEKLTLCFYFRRVTEKKPTGLVTFTRQSLDPPPEWERCEKPLTRLH 719  
QY 712 VTIEGTIEGNRGMQLQVDFANRFVGGVGTGAGLVQEEIRFLINPELIVSRLEFTEVLDHNE 771  
Db 720 VTIEGTIEGNRGMQLQVDFANRFVGGVGTGAGLVQEEIRFLINPELIVSRLEFTEVLDHNE 779

QY 772 CLITGTEQYSEYTGAEYRWARSHEDGSEKDDQRRCTEIVADALHFRYLDQFVPE 831  
Db 780 CLITGTEQYSEYTGAEYRWSRSHEDGSRDDQRRCTEIVADALHFRYLDQFVPE 839  
QY 832 KVRRELKAYCGFLRPGVPSNL SAVATGNWCGAGGDDARLKALIQILAAAAAERDVVY 891  
Db 840 KVRRELKAYCGFLRPGVSSNL SAVATGNWCGAGGDDARLKALIQILAAAAAERDVVY 899  
QY 892 FTFGDSELMRDIYSMHIFLTERKLDVGKVVYKLLRLRYNNEECRCNCSTPGPDIKLYPFIYHA 951  
Db 900 FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNNEECRCNCSTPGPDIKLYPFIYHA 959  
QY 952 VESSAETDMPGQKAGT 968  
Db 960 VESCAETADHSGQRTGT 976  
RESULT 6  
Q7Z742 PRELIMINARY; PRT; 976 AA.  
AC Q7Z742;  
DT 01-OCT-2003 (TremBLrel. 25, Created)  
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=2388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC052966; AAH52966.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8BE1826C4A74A CRC64;

Query Match 86.2%; Score 4451; DB 4; Length 976;  
Best Local Similarity 86.5%; Pred. No. 0;  
Matches 845; Conservative 49; Mismatches 73; Indels 10; Gaps 6;  
QY 1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSPGRRRLVDPKAPVQFRVPPSSPACVS 59  
Db 1 MNAGPGCEPCTKRPWGAA-TTSPAASDARSFPSPRRRLVDPKDAHVQFRVPPSSPACVP 59  
QY 60 GRAGPHRGATSFVFKQKTTTWMOTKGPXTAESE---SKENNTRIDSMSSVQKDNFY 116  
Db 60 GRAGQHRGATSLSLVFKQKTTTWMOTKGIKTAESSELDSEKNNTRIDSMSSVQKDNFY 119

QY 117 PHKVEKLENVQLNLDKSPTEKSSQYLNQOQTASVCKWQNEGKHAEQLLASEPPAGTLP 176  
Db 120 QHNVEKLENVQLSLDKSPTEKSTQYLNQHQTAAAMCKWQNEGKHTEQLLESEPQTTLVP 179  
QY 177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQAR---SNCKCSG 233  
Db 180 EQFSNANIDRSQNDHSDTDSEENRDNQOFLTTVKLANAKQTTEDEQAREAKSHQKCSK 239  
QY 234 SRQSVKDCCTGCQEEVDVLPESPLSDVGAEDTGTGPKNDNKLGTQESSLSPPFKEKESE 293  
Db 240 SCDPGEDCASCQDEIDVVPESPLSDVGSSEVGTGPKNDNKLTRQESCLGNSPPFEKESE 299  
QY 294 PESPMVDNRSNCQDSEADEETS PVFEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352  
Db 300 PESPMVDNRSNCQDSEADEETS PVFEQDDGSSSSQTANKPSRFQARDADIEFRKRYST 359  
QY 353 KGSEVRLHFQFE-GENNAGTSDLNAPSGNSSSLNVECRSSKQHGKRDSKI THFMRIK 411  
Db 360 KGSEVRLHFQFEGGESRTGMNDLNAPGNISSLNVECRNSKQHGKDKSKI THFMRLPK 419  
QY 412 SEDRRKEQCEVHRHQRTERKIPKYIPPNLPPEKKWLTGPIEMRMKPRCGIHLPSLRPSAS 471  
Db 420 AEDRRKEQWETKHQRTERKIPKYVPPHLSDPKRWLTGPIEMRRMPCGIRLPLRPSAN 479  
QY 472 HTVTVRVDLLRAGEVPKPPPTHYKLDWKNHVMKPCSEQNLYPVEDENGERTAGSRWELI 531  
Db 480 HTVTIRVDLLRAGEVPKPPPTHYKLDWKNHVMKPCSEQNLYPVEDENGERTAGSRWELI 539  
QY 532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFALTALDFWDKVLLEAEAAHQLYQSILPDMVK 591  
Db 540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFALTALDFWDKVLLEAEAAHQLYQSILPDMVK 599  
QY 592 IALCLPNICTQPIPLLKQKNHVSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 651  
Db 600 IALCLPNICTQPIPLLKQKNHVSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659  
QY 652 NFNRLFEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFTRQSLDFPEWERCEKPLTRLH 711  
Db 660 NFNRLFEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFTRQSLDFPEWERCEKPLTRLH 719  
QY 712 VTYEGTIEGNRGMQLQVDFANRFVGGVTGAGLVQEEIRFLINPELIVSRFLTVEVDHNE 771  
Db 720 VTYEGTIEGNRGMQLQVDFANRFVGGVTGAGLVQEEIRFLINPELIVSRFLTVEVDHNE 779  
QY 772 CLITGTEQYSEYTGAEYRWARSHEDGSEKDDQRRCTEIVADALHFRYLDQFVPE 831  
Db 780 CLITGTEQYSEYTGAEYRWSRSHEDGSRDDQRRCTEIVADALHFRYLDQFVPE 839  
QY 832 KVRRELKAYCGFLRPGVPSNL SAVATGNWCGAGGDDARLKALIQILAAAAAERDVVY 891  
Db 840 KVRRELKAYCGFLRPGVSSNL SAVATGNWCGAGGDDARLKALIQILAAAAAERDVVY 899  
QY 892 FTFGDSELMRDIYSMHIFLTERKLDVGKVVYKLLRLRYNNEECRCNCSTPGPDIKLYPFIYHA 951  
Db 900 FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNNEECRCNCSTPGPDIKLYPFIYHA 959  
QY 952 VESSAETDMPGQKAGT 968  
Db 960 VESCAETADHSGQRTGT 976  
RESULT 7  
Q9Y4W7  
ID Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
AC Q9Y4W7;  
DT 01-NOV-1999 (TremBLrel. 12, Created)  
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)  
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



|    |  |  |  |  |  |
|----|--|--|--|--|--|
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.           |  |  |  |  |
| OX | NCBI_TaxID=9606;   |  |  |  |  |
| RN | [1]  |  |  |  |  |
| RP | SEQUENCE FROM N.A.   |  |  |  |  |
| RX | MEDLINE=99380098; PubMed=10449915;                                   |  |  |  |  |
| RA | Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;                    |  |  |  |  |
| RT | "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to    |  |  |  |  |
| RT | human chromosome 10q11.23 and mouse chromosome 14B by in situ        |  |  |  |  |
| RT | hybridization."  |  |  |  |  |
| RL | Cytogenet. Cell Genet. 85:269-270 (1999).                            |  |  |  |  |
| DR | EMBL; AF005043; AAB61614.1; -.                                       |  |  |  |  |
| DR | Genew; HGNC:8605; PARG.  |  |  |  |  |
| DR | GO; GO:0005737; C:cytoplasm; TAS.                                    |  |  |  |  |
| DR | GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.     |  |  |  |  |
| DR | InterPro; IPR007724; PARG.   |  |  |  |  |
| DR | Pfam; PF05028; PARG; 1.  |  |  |  |  |
| KW | Hydrolase.   |  |  |  |  |
| SQ | SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;                  |  |  |  |  |
|    | Query Match 85.1%; Score 4395; DB 4; Length 976;                     |  |  |  |  |
|    | Best Local Similarity 85.7%; Pred. No. 0;                            |  |  |  |  |
|    | Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;      |  |  |  |  |
| QY | 1 MSAGPGWEPCTKA-RWGAGTSAPTASDSRSFPGRRVLDPKDAPVQFRVPPSPACVS 59        |  |  |  |  |
| Db | 1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFSPRRVLDPKDAHVQFRVPPSPACVP 59       |  |  |  |  |
| QY | 60 GRAGPHRGNATSFVFKQKTTITWMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY 116   |  |  |  |  |
| Db | 60 GQAGQHRSATSLVFKQKTTITWMDTKGPKTAESESLDSKNNNTRIESMMSSVQKDNFY 119    |  |  |  |  |
| QY | 117 PHKVEKLENVLPQLNLDKSPTEKSSQYLNQOQTASVCKWQNEGKHAEQLLASEPPAGTLP 176 |  |  |  |  |
| Db | 120 QHNVEKLVNVSQSLDKSLTEKSTQYLNQHTAAMCKWQNEGKHTEQLLSEPTVTLVP 179     |  |  |  |  |
| QY | 177 KOLSANIGQSPHTDDHSDTDHEEDRNQOFLTPIKLANTKPTVGDGQAR---SNCKCSG 233   |  |  |  |  |
| Db | 180 EQFSNANIDRSPQNDHSDTDSEENRDNQOFLTITVKLANAKQTTEDEHAREAKSHQKCSK 239 |  |  |  |  |
| QY | 234 SRQSVKDCGCGQEEVDVLPESPLSDVCAEDIGTGPKNDKLTGQESSLGDSPPFEKES 293    |  |  |  |  |
| Db | 240 SCHPGEDCASCQDEIDVVPKSPLSVGSSEVDVGTGSKNDKLTIRQESCLGNSPPFEK 299    |  |  |  |  |
| QY | 294 PESPMVDVNSRNSQDSEADETSPVDEBQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352   |  |  |  |  |
| Db | 300 PESPMVDVNSKNSQDSEADETSPGDFEQEDGSSSQTANKPSRFQARDADIEFRKRY 359     |  |  |  |  |
| QY | 353 KGSEVRLHQPGE-GENNAGTSDLNAPKPSGNSSLNVECRSSKHQGRKDSKITDHEMRIS 411  |  |  |  |  |
| Db | 360 KGSEVRLHQPGEESRTGMNDLNAPKPSGNSSLNVECRSSKHQGRKDSKITDHLRLPK 419    |  |  |  |  |
| QY | 412 SEDRRKEQCEVRHQRTERKIPKYPNLPPEKKWLTGTPIEEMRMKPRCGIHLPSLRPSAS 471  |  |  |  |  |
| Db | 420 AEDRRKEQWETKHQRTERKIPKYPVPHLSGDKWLTGTPIEEMRMKPRCGIRLPLRPSAN 479  |  |  |  |  |
| QY | 472 HTVTIRVDLLRAGEVPKPPPTHYKOLWDNKHVMKPCSEQNLYPVEDENGERTAGSRWELI 531 |  |  |  |  |
| Db | 480 HTVTIRVDLLRAGEVPKPPPTHYKOLWDNKHVMKPCSEQNLYPVEDENGERTAGSRWELI 539 |  |  |  |  |
| QY | 532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFDALVDFWDKVLAEAAQHLYSILPDMVK 591   |  |  |  |  |
| Db | 540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFDALVDFWDKVLAEAAQHLYSILPDMVK 599   |  |  |  |  |
| QY | 592 IALCLPNICTQPIPLLKQWNSHTVMSQEQIASILANAFCTFPRRNAMKMSSEYSSYPDI 651  |  |  |  |  |
| Db | 600 IALCLPNICTQPIPLLKQWNSHTVMSQEQIASILANAFCTFPRRNAMKMSSEYSSYPDI 659  |  |  |  |  |
| QY | 652 NFNRLFEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFTRQSLDFFPEWERCEKPLTRLH 711  |  |  |  |  |
| Db | 660 NFNRLFEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFTRQSLDFFPEWERCEKPLTRLH 719  |  |  |  |  |
| QY | 712 VTTEGTIEGNRGMQLQVDFANRFVGGVGTAGLVQSEIRFLINPELIVSRILFTEVLHDNE 771 |  |  |  |  |

|    |   |  |  |  |  |
|----|---|--|--|--|--|
| Db | 720 VTTEGTIEENGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNE 779   |  |  |  |  |
| QY | 772 CLIIITGTEQYSEYGYAETRYRWSHEDGSEKODWQRRCTEIVDALHFRRYLDQFVPE 831     |  |  |  |  |
| Db | 780 CLIIITGTEQYSEYGYAETRYRWSHEDGSEKODWQRRCTEIVDALHFRRYLDQFVPE 839     |  |  |  |  |
| QY | 832 KVRRELKAYCGFLRPGVPSSENLAVATGNWCGAGFGGDARLKALIQILAAAAAERDVVY 891   |  |  |  |  |
| Db | 840 KMRRELKAYCGFLRPGVPSSENLAVATGNWCGAGFGGDARLKALIQILAAAAAERDVVY 899   |  |  |  |  |
| QY | 892 FTFGDSELMDIYSMHTFLTERKLDVGKVVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHA 951 |  |  |  |  |
| Db | 900 FTFGDSELMDIYSMHTFLTERKLDVGKVVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHA 959 |  |  |  |  |
| QY | 952 VESSAETTMPCQKAGT 968  |  |  |  |  |
| Db | 960 VESCAETADHSGORTGT 976   |  |  |  |  |
|    | RESULT 8  |  |  |  |  |
| ID | O02776 PRELIMINARY; PRT; 977 AA.                                      |  |  |  |  |
| AC | O02776;   |  |  |  |  |
| DT | 01-JUL-1997 (Tremblrel. 04, Created)                                  |  |  |  |  |
| DT | 01-JUL-1997 (Tremblrel. 04, Last sequence update)                     |  |  |  |  |
| DT | 01-JUN-2003 (Tremblrel. 24, Last annotation update)                   |  |  |  |  |
| DE | Poly(ADP-ribose) glycohydrolase.                                      |  |  |  |  |
| GN | BPARG.  |  |  |  |  |
| OS | Bos taurus (Bovine).  |  |  |  |  |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;     |  |  |  |  |
| OC | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;    |  |  |  |  |
| OC | Bovidae; Bovinae; Bos.  |  |  |  |  |
| OX | NCBI_TaxID=9913;  |  |  |  |  |
| RN | [1]   |  |  |  |  |
| RP | SEQUENCE FROM N.A.  |  |  |  |  |
| RX | MEDLINE=97277328; PubMed=9115250;                                     |  |  |  |  |
| RA | Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;         |  |  |  |  |
| RT | "Isolation and characterization of the cDNA encoding bovine poly(ADP- |  |  |  |  |
| RT | ribose) glycohydrolase."  |  |  |  |  |
| RL | J. Biol. Chem. 272:11895-11901 (1997).                                |  |  |  |  |
| DR | EMBL; U78975; AAB53370.1; -.  |  |  |  |  |
| DR | GO; GO:0016787; F:hydrolase activity; IEA.                            |  |  |  |  |
| DR | InterPro; IPR007724; PARG.  |  |  |  |  |
| DR | Pfam; PF05028; PARG; 1.   |  |  |  |  |
| KW | Hydrolase.  |  |  |  |  |
| SQ | SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;                   |  |  |  |  |
|    | Query Match 83.9%; Score 4332.5; DB 6; Length 977;                    |  |  |  |  |
|    | Best Local Similarity 84.0%; Pred. No. 8.8e-314;                      |  |  |  |  |
|    | Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;        |  |  |  |  |
| QY | 1 MSAGPGWEPCTK-ARWGAGTSAPTASDSRSFPGRRVLDPKDAPVQFRVPPSPACVS 59         |  |  |  |  |
| Db | 1 MSAGPGCEPCTKRPRWDAATSPFAASDARSFPGRRVLDKADAPVQFRVPPSSGCGAL 60        |  |  |  |  |
| QY | 60 GRAGPHRGNATSFVFKQKTTITWMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY 116    |  |  |  |  |
| Db | 61 GRAGQHRSATSLVFKQKTTITWMDTKGPKTAESESLHSKNNNTREESMMSSVQKDNFY 120     |  |  |  |  |
| QY | 117 PHKVEKLENVLPQLNLDKSPTEKSSQYLNQOQTASVCKWQNEGKHAEQLLASEPPAGTLP 176  |  |  |  |  |
| Db | 121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAAMCKWQNEGKHTEQLLSEPPAVTLVP 180  |  |  |  |  |
| QY | 177 KOLSANIGQSPHTDDHSDTDHEEDRNQOFLTPIKLANTKPTVGDGQ---ARSNCCKSG 233    |  |  |  |  |
| Db | 181 EQFSNANVDSQSPKDDHSDTNSESRDNQOFLTITVKLANAKQITMEDEQGREARSHQKCGK 240 |  |  |  |  |
| QY | 234 SRQSVKDCGCGQEEVDVLPESPLSDVCAEDIGTGPKNDKLTGQESSLGDSPPFEKES 293     |  |  |  |  |
| Db | 241 ACHPAEACAGCQEEETDVVSESPSLDTGSEDVGTGLKNANRLNRFQESSLGNSPPFEKES 300  |  |  |  |  |
| QY | 294 PESPMVDVNSRNSQDSEADETSPVDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352     |  |  |  |  |





|  |  |  |     |
|--|--|--|-----|
| QY   | 644  | EYSSYPDINFNRLFEGRSSRRKPEKLTLCFYFRRV--TEKK-----PTGLVTFTRQS--LE  | 695 |
| Db   | 272  | EYSTFPDINFNRLYQSTGTPAVLEKLCIMHYFRRVCPTERDASNVPTGVTFRRSGLPE     | 331 |
| QY   | 696  | DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLOVDFAANRFVGGVGTGAGLVQEEIRFLI  | 753 |
| Db   | 332  | HLIDWSQSAAPLGDVPLHVDAEGTIEDEGILLQVDFANKYLGGLGGVGLGHGCVQEEIRFVI | 391 |
| QY   | 754  | NPELIVSRLFTEVLDHNECLIITGTQYSEYTYGAETYRWARSHEDSEKDDWQRRCTEI     | 813 |
| Db   | 392  | CPELLVGKLFTECLRPFEALVMLGAERYSNYTYAGSFEWSGNFEDSTPRDSSGRRQTAI    | 451 |
| QY   | 814  | VAIDALHFRRYLDQFVPEKVRRELKAYCGFLR-----PGVPSNLSAVATGNWCGGAF      | 867 |
| Db   | 452  | VAIDALHFAQSHHGYREDLMERELNKAYIGFVHWMVTPPPG-----VATGNWCGGAF      | 503 |
| QY   | 868  | GGDARLKALIQILAAAAAARDVVYFTFGDSELMRDIYSMHTFLTTERKLDVGKVKLLRY    | 927 |
| Db   | 504  | GGDSYLKALLQLMVCAQLGRPLAYYTFGNVFEPRDDFHEMWLLFRNDGTTVQQLWS-ILRS  | 562 |
| QY   | 928  | YNEECRNCSTPGP-----DIKLYPFIYHAVESSAETTDMPGQKA 966               |     |
| Db   | 563  | YSRLIKEKSSKEPRENKASKKLYDFI---KEELKKVRDVPGEA 604                |     |
| RESULT 10  |  |  |     |
| ID   | Q960N8   | PRELIMINARY; PRT; 723 AA.                                      |     |
| AC   | Q960N8;  |  |     |
| DT   | 01-DEC-2001 (TREMBlrel. 19, Created)                                 |  |     |
| DT   | 01-DEC-2001 (TREMBlrel. 19, Last sequence update)                    |  |     |
| DT   | 01-JUN-2003 (TREMBlrel. 24, Last annotation update)                  |  |     |
| DE   | LD42380p.  |  |     |
| GN   | PARG OR EG:114E2.1 OR CG2864.  |  |     |
| OS   | Drosophila melanogaster (Fruit fly).                                 |  |     |
| OC   | Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;                  |  |     |
| OC   | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;           |  |     |
| OC   | Ephydroidea; Drosophilidae; Drosophila.                              |  |     |
| OX   | NCBI_TaxID=7227;   |  |     |
| RN   | [1]  |  |     |
| RP   | SEQUENCE FROM N.A.   |  |     |
| RC   | STRAIN=Berkeley;   |  |     |
| RA   | Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,        |  |     |
| RA   | Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,    |  |     |
| RA   | Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,    |  |     |
| RA   | Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,    |  |     |
| RA   | Yu C., Lewis S.E., Rubin G.M., Celniker S.;                          |  |     |
| RL   | Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.              |  |     |
| DR   | EMBL; AY051955; AAK93379.1; -.                                       |  |     |
| DR   | FlyBase; FBgn0023216; PARG.  |  |     |
| DR   | InterPro; IPR007724; PARG.   |  |     |
| DR   | Pfam; PF05028; PARG; 1.  |  |     |
| SQ   | SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;                   |  |     |
| Query Match  |  |  |     |
| Best Local Similarity 20.7%; Score 1069.5; DB 5; Length 723;       |  |  |     |
| Matches 243; Conservative 96; Mismatches 194; Indels 51; Gaps 15;  |  |  |     |
| QY   | 409  | ISKSDRRKEQCEVRHQRTERKIPKYPNLP--EKKWLGTPIEMRKMPCRGHLPSSL        | 466 |
| Db   | 1  | MSKSPDGGISEIETEE-----PENLANSLDSDSWRGVSMEAIHR-NRQPFLENL         | 49  |
| QY   | 467  | RP-SASHTVTVRVDLLRAGEVPKPFPHYKLDWKNHKVMPQSEQNLYPVEDENGERTAG     | 525 |
| Db   | 50   | PPVTAGNLHRVMYQLPIRETPPRPYKSPGK--WDSEHVRLPAPESKYPRENPDGSTTID    | 107 |
| QY   | 526  | SRWELIQTALLNKFRPNQLKDALKNVAYSKKWDFTALVDFWDKVLLEAEQAHLQYQSI     | 585 |
| Db   | 108  | FRWEMIERALLQPIKTCCELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFEDL   | 167 |
| QY   | 586  | LPMVMKIALCLPNICTQPIPLLKQNMHVSVTMSQEQIASLLANAFCTFPRRNA-YMKSE    | 644 |
| Query Match  |  |  |     |
| Best Local Similarity 41.6%; Pred. No. 7.6e-71;                    |  |  |     |
| Matches 243; Conservative 96; Mismatches 194; Indels 51; Gaps 15;  |  |  |     |
| QY   | 644  | EYSSYPDINFNRLFEGRSSRRKPEKLTLCFYFRRV--TEKK-----PTGLVTFTRQS--LE  | 695 |
| Db   | 272  | EYSTFPDINFNRLYQSTGTPAVLEKLCIMHYFRRVCPTERDASNVPTGVTFRRSGLPE     | 331 |
| QY   | 696  | DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLOVDFAANRFVGGVGTGAGLVQEEIRFLI  | 753 |
| Db   | 332  | HLIDWSQSAAPLGDVPLHVDAEGTIEDEGILLQVDFANKYLGGLGGVGLGHGCVQEEIRFVI | 391 |
| QY   | 754  | NPELIVSRLFTEVLDHNECLIITGTQYSEYTYGAETYRWARSHEDSEKDDWQRRCTEI     | 813 |
| Db   | 392  | CPELLVGKLFTECLRPFEALVMLGAERYSNYTYAGSFEWSGNFEDSTPRDSSGRRQTAI    | 451 |
| QY   | 814  | VAIDALHFRRYLDQFVPEKVRRELKAYCGFLR-----PGVPSNLSAVATGNWCGGAF      | 867 |
| Db   | 452  | VAIDALHFAQSHHGYREDLMERELNKAYIGFVHWMVTPPPG-----VATGNWCGGAF      | 503 |
| QY   | 868  | GGDARLKALIQILAAAAAARDVVYFTFGDSELMRDIYSMHTFLTTERKLDVGKVKLLRY    | 927 |
| Db   | 504  | GGDSYLKALLQLMVCAQLGRPLAYYTFGNVFEPRDDFHEMWLLFRNDGTTVQQLWS-ILRS  | 562 |
| QY   | 928  | YNEECRNCSTPGP-----DIKLYPFIYHAVESSAETTDMPGQKA 966               |     |
| Db   | 563  | YSRLIKEKSSKEPRENKASKKLYDFI---KEELKKVRDVPGEA 604                |     |
| RESULT 11  |  |  |     |
| ID   | Q9SKB3   | PRELIMINARY; PRT; 548 AA.                                      |     |
| AC   | Q9SKB3; Q94ET7;  |  |     |
| DT   | 01-MAY-2000 (TREMBlrel. 13, Created)                                 |  |     |
| DT   | 01-JUN-2002 (TREMBlrel. 21, Last sequence update)                    |  |     |
| DT   | 01-OCT-2003 (TREMBlrel. 25, Last annotation update)                  |  |     |
| DE   | Putative poly (ADP-ribose) glycohydrolase.                           |  |     |
| GN   | AT2G31870 OR TEJ.  |  |     |
| OS   | Arabidopsis thaliana (Mouse-ear cress).                              |  |     |
| OC   | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;   |  |     |
| OC   | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; |  |     |
| OC   | eurosid II; Brassicales; Brassicaceae; Arabidopsis.                  |  |     |
| OX   | NCBI_TaxID=3702;   |  |     |
| RN   | [1]  |  |     |
| RP   | SEQUENCE FROM N.A.   |  |     |
| RC   | STRAIN=cv. Columbia;   |  |     |
| RA   | Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,       |  |     |
| RA   | Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I., |  |     |
| RA   | Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,      |  |     |
| RA   | Fraser C.M., Venter J.C.;  |  |     |
| RL   | Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.              |  |     |
| RN   | [2]  |  |     |
| RP   | SEQUENCE FROM N.A.   |  |     |
| RA   | Panda S., Poirier G.G., Kay S.A.;                                    |  |     |
| RT   | "TEJ defines a role for poly-ADP-ribosylation in establishing period |  |     |
| RT   | length of the Arabidopsis circadian oscillator."                     |  |     |
| RL   | Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.              |  |     |
| DR   | EMBL; AC006533; AAD32285.2; -.                                       |  |     |
| DR   | EMBL; AF394690; AAK72256.1; -.                                       |  |     |
| DR   | PIR; B84726; B84726.   |  |     |
| DR   | GO; GO:0016787; F:hydrolase activity; IEA.                           |  |     |
| DR   | InterPro; IPR007724; PARG.   |  |     |
| DR   | Pfam; PF05028; PARG; 1.  |  |     |
| KW   | Hydrolase.   |  |     |
| SQ   | SEQUENCE 548 AA; 62169 MW; F1A79FDA157C329 CRC64;                    |  |     |
| Query Match  |  |  |     |
| Best Local Similarity 11.0%; Score 568; DB 10; Length 548;         |  |  |     |
| Matches 147; Conservative 61; Mismatches 139; Indels 104; Gaps 11; |  |  |     |
| QY   | 569  | WDKVLLEAEQAHLQYQSILPDMVKIALCLPNI-----CTQPIPLLKQNMHNS           | 614 |
| Db   | 79   | FDELIDKEKSRWFDEIIPALASLLQLQFSPSLLEVHFQFNADNIVSGIKTGLRLNSQQAGI  | 138 |
| QY   | 615  | VTMSQEQIASLLANAFCTFPRRNAKMKSEYSSYPDINFNRLFE---GRSSRKPEKLKT     | 670 |

```
Db      139 VFLSQELIGALLACSFCLFPDDNRGAK-----HLPVINFDHLPASLYISYSQESKIRC 194
QY      671 LFCYFRRVTEKKTGLVTFTRQ--SLEDPEWERCEKPLTRLHLVTVYEGTIEGNRGLQV 728
Db      195 IMHYFERFCSCVPIGIVSFERKITAAPDADFWSKSDVSLCAFKVHSFGLIEDQPDNALEV 254
QY      729 DFANRFVGGVGTAGLVQEEIRFLINPELVSRFLTFEVLHDNECLITGTQYSEYTYA 788
Db      255 DFANKYLGGSLSRGCVQEEIRFMINPELIAGMLFLPRMDDNEAIEIVGAERFSCYTGYA 314
QY      789 ETYRWARSHEDGSEKDDWQRRCTEIVAIDAL-----HFRRYLDQFVPEKVRRELNKAYC 842
Db      315 SSFRFAGEYIDKKAMDPEKRRTRIVAIDALCTPKMRHFKDIC-----LLREINKALC 367
QY      843 GFL-----RPG-----VPSEN 853
Db      368 GFLNCSKAWEHQNFMDEGDNEIQLVNRGRDSSLRTTETTASHRTPLNDVEMNREKPA 427
QY      854 L-----SAVATGNWGGAGFGGDARLKALIQILAAAAAERDVV-YFTFGD 896
Db      428 LIRDFYVEGVNDHEDDGVATGNWGGVFGGDEPKATIQWLAAASQTRRPFISYYTFG- 486
QY      897 SELMRDIYSMHTFLTERKLDVGKVKLLRLRY 927
Db      487 VEALRNLDQVTKWILSHKWTVDGLWNMMLEY 517
RESULT 12
Q8VYAI ID Q8VYAI PRELIMINARY; PRT; 522 AA.
AC Q8VYAI;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Putative poly(ADP-ribose) glycohydrolase.
GN AT2G31870.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY072330; AAL61937.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;
Query Match 10.2%; Score 525; DB 10; Length 522;
Best Local Similarity 30.5%; Pred. No. 1.7e-30;
Matches 132; Conservative 82; Mismatches 139; Indels 80; Gaps 11;
QY 568 FWDKVLAEAAQHLQSIILPDMVKIALCLPNICTQPIPLKQKMNH----- 613
Db 80 FFDKKSREESANFFGEVVPALCRLQLPSMLEKHY----QKADHVLGDKVSGRLRLGP 135
QY 614 ----SVTMSQEQIASLLANAFCTFPRNNAKMKSEYSSYPDINFRNLF-----GRSSRKP 665
Db 136 QEAGIVLLSQELIAALLACSFCLFPEVDRSLK----NLQGINFSGLFSFPYMRHCTKQE 191
QY 666 EKLKTLFCYFRRVTEKKTGLVTFTRQSLR-----DFPE----WERCEKPLTRLHVTY 714
Db 192 NKIKCLIHVFGICRWMPGTGVSFERKILPLEYHPHFVSYPKADSWANSVTPLCSIEIHT 251
```

```
QY      715 EGTIEGNRGMQLQVDFANRFVGGVGTAGLVQEEIRFLINPELVSRFLTFEVLHDNECLI 774
Db      252 SGAIEDQPCAEALEVDFADEYFGGLTSLSYDTLQEEIRFVINPELIAGMIFLPRMDANEAE 311
QY      775 ITGTEQYSEYTYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVR 834
Db      312 IVGVERFSYTYGSPSFQYAGDYTDNKDLDFRRRKTRVIAIDAMP-DPGMGQYKLDALI 370
QY      835 RELNKAYCGFLR-----PGVPSENLSA-----VA 858
Db      371 REVNKAFSGYMHQCKYNIDVKHDPEASSSHVPLTSDSASQVIESSHRWICIDHEEKIGVA 430
QY      859 TGNWGGAGFGGDARLKALIQILAAAAAERDVV-YFTFGDSELMRDIYSMHTFLTERKLDV 917
Db      431 TGNWGGVFGGDEPKLIMLQWLAIQSQGRPFMSYYTFG-LQALQNLNQVIEMVALQEMTV 489
QY      918 GKVKLLRLRYNE 930
Db      490 GDLWKKLVEYSSE 502
RESULT 13
Q9N5L4 ID Q9N5L4 PRELIMINARY; PRT; 485 AA.
AC Q9N5L4;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme
DE 4)
GN H23L24.5 OR PME-4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RC "Molecular characterization of poly ADP-ribose glycohydrolases in
RC Caenorhabditis elegans.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006662; AAF39896.2; -.
DR EMBL; AFS48468; AAN40699.1; -.
DR WormPep; H23L24.5; CE32685.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hypothetical protein.
SQ SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;
Query Match 8.9%; Score 458; DB 5; Length 485;
Best Local Similarity 29.3%; Pred. No. 1.5e-25;
Matches 142; Conservative 84; Mismatches 160; Indels 98; Gaps 18;
```



|    |     |   |     |
|----|-----|---|-----|
| QY | 528 | WELIQTALLNKETR-PQNLKDAILK--YNVAYS-----KKWDF TALVDF            | 568 |
| Db | 46  | FELLETVGSQQWRNCQNLFNEYLKYTKNGGYSQFEDLLFKIWGYSEKERFDLPALKSF    | 105 |
| QY | 569 | WDKVLE---EAEAQHLYSILPDMVKIALCLPNICTQPIPLLKQKN-----HsvTMSQ     | 619 |
| Db | 106 | YRKMSEIVGEDE-----VLEKLARLVRIITKSAC---EVLPEKIYRLVGDIESATFSH    | 154 |
| QY | 620 | EQIASLLANAFECTFPRRNAKMKSEYSSYPDINFNRLFEGRSRKPPEKLTLCFYFRRVT   | 679 |
| Db | 155 | IQCASLIAMWFFSDTPR-----LSFIILQKTTCVAVEKLLKFLTIFYFDKMS          | 200 |
| QY | 680 | EKKPTGLVTFT-----ROSLEDFFPEWE-RCEKPLTRLHVTYEGTIEGNRGRMLQVDFAN  | 732 |
| Db | 201 | IDPPIGAVSFRKMRIITHKQYLEN--WKLRETNLLPDVQVFDKMSIETAL-CTQIDFAN   | 256 |
| QY | 733 | RFVGGGVGTAGLVQEEIRFLINPELIVSRFLTTEVLHDNECLIIITGTQYSEYTYAETVR  | 792 |
| Db | 257 | KRLGGGVULKGGAVQEEIRFMMCPENMVAILLNDVTQDLEAISIVGAYVFSSYTGYSNTLK | 316 |
| QY | 793 | WAR-----SHEDGSEKDDWQRCCTEIVDAIDALH-----FRRYLDQFVPEKVRRELNKAY  | 841 |
| Db | 317 | WAKITPKHSAQNNSFRDQFGRLQTEVVAIDAVRNAGTPLECLLNQLTTEKLTREVRKAA   | 376 |
| QY | 842 | CGFLRPG-----VPSENLSAVATGNWGCAGGDDARLKALIQILAAAAERDVVYFTFGD    | 896 |
| Db | 377 | IGFLSAGDGFSKIP-----VVSQWGWGCGAFRGNKPKLFQIACGISDRPLQFCTFGD     | 430 |
| QY | 897 | SELMRDIYSMHTFLTERKLDVGKVVYKLLRLRYNEECRNCSTPGPDIKL-YPFYIHAVESS | 955 |
| Db | 431 | TELAKKCEEMMTLFRNNNVRTGQLFLII-----NSIGPPLNYSQVVFDAIRAK         | 479 |
| QY | 956 | AETT 959  |     |
| Db | 480 | INST 483  |     |

RESULT 14

Q867X0 PRELIMINARY; PRT; 781 AA.

AC Q867X0;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3 protein) (Corresponding sequence F20C5.1a).

GN F20C5.1 OR PME-3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Gagnon S.N., Hardy I., Desnoyers S.;

RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode Caenorhabditis elegans.";

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";

RL Science 282:2012-2018(1998).

RN [3]

RP SEQUENCE FROM N.A.

RA Matthews P.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY185493; AAC26316.1; -.

DR EMBL; Z68161; CAD89735.1; -.

DR WormPep; F20C5.1a; CE33775.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

| SQ | SEQUENCE              | 781 AA;   | 89273 MW;        | 30447D06605CED0A | CRC64;      |
|----|-----------------------|---|------------------|------------------|-------------|
|    | Query Match           | 8.7%;   | Score 451;       | DB 5;            | Length 781; |
|    | Best local Similarity | 22.9%;  | Pred. No. 1e-24; |                  |             |
|    | Matches 197;          | Conservative 137;   | Mismatches 297;  | Indels 230;      | Gaps 37;    |
| QY | 199                   | BEDRDNQQLTPIKLANTKPTVGDGQARSNCCKSGSRQSVKDOCTGCQQEVDVLPESPLS   | 258              |                  |             |
| DB | 14                    | QDEKDYEDYV-GVGFAHQVPTM--KRRKLTGHTGNTESKEDPE--EPKSRDVFVSSQSS   | 67               |                  |             |
| QY | 259                   | DVGAEDIGTGPK-----NDNKLTGQESSLGD---SPPFEXES                    | 292              |                  |             |
| DB | 68                    | DESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLDNHKSTEPMEEDV  | 127              |                  |             |
| QY | 293                   | EPESPMVDVNSRNSQDSE-----ADETSPVF-DEQD-----                     | 324              |                  |             |
| DB | 128                   | NNKSNIDV--AINSEDEDELVLEENKEMRDGEQVQODLFAADDQELIEYPGIMKDTTTL   | 185              |                  |             |
| QY | 325                   | ---DRSSQTANKLSSCQAREADGLRKRYLTKGSEVRLHFQEGENNAGTSDLNAKPSGN    | 381              |                  |             |
| DB | 186                   | DITDSEVETAQMEMIEETADS-----TFVGEDSK-----NQROSGT                | 223              |                  |             |
| QY | 382                   | S-----SSLNVECRSSQHGRKDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIPKY     | 434              |                  |             |
| DB | 224                   | TSDEVDADSQINLATKYR-----TSSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAF    | 277              |                  |             |
| QY | 435                   | IPPNLPEKKWLGTPIEMRKMPRCGIHLPSLRPSASHVTVRVD---LLRAGEVPKPEP     | 491              |                  |             |
| DB | 278                   | TEGNL-----TLQPDNL-----KVPDPNRYRYCTIPN-PP                      | 306              |                  |             |
| QY | 492                   | THYKDL-WDNKH---VKMPCSEONLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKD  | 547              |                  |             |
| DB | 307                   | ASQGLREDNRYGPKIVLP-----QRW-----REFDSRGRRRD                    | 339              |                  |             |
| QY | 548                   | AILKYN---VAYSKKWDFTA-----LVDFWD---KVLSEEAQAHLYSILPDMV---      | 590              |                  |             |
| DB | 340                   | SYFYFKRLDGLYKCYKTYGTFYMFVGLLHNMWEFPDITYKLPALEMYIKEMSELVGREE   | 399              |                  |             |
| QY | 591                   | -----KIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASILLANAFFCTFPRRNAKM     | 641              |                  |             |
| DB | 400                   | VLEKFAVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCAALVARMFF-----ARP     | 451              |                  |             |
| QY | 642                   | KSEYSSYPDINFNRLFEGRSRKPPEKIKTLFCYFRRRVTEKKPTGLVTF--TRQSLDF-P  | 698              |                  |             |
| DB | 452                   | DSPFS-----FCRILSSKSIQVEKLFLETTYFDKMSMDPPDGAVSFRLTKMDKDTFNE    | 505              |                  |             |
| QY | 699                   | EW-ERCEKPLTRLHVTEGTIEGNGRGMLOVDFAANRFVGGVGTGAGLVQEEIRFLINPEL  | 757              |                  |             |
| DB | 506                   | EWKDKLRSLPEVEFFDEMLEDIAL-CTQVDFANEHLGGVVLNHGVSQEEIRFLMCPM     | 564              |                  |             |
| QY | 758                   | IVSRLFTEVLDHNECLIIITGTEQYSEYTYGAETYRWA-----RSHEDGSE-KDDWQRRCT | 811              |                  |             |
| DB | 565                   | MVGMLLCEKMKQLEAISIVGAYVPSSYTYGYHTLTKWAELOPNHSPQNTNEFRDRFGLRLV | 624              |                  |             |
| QY | 812                   | EIVAIIDALHFR-----YLDQFVPEKVRRELNKAYCGFLRPGVPSNLSAVATGNWCGGA   | 866              |                  |             |
| DB | 625                   | ETIAIDAILFKGSLDCQTEQLNKANIIREMKKASIGFMSQGPKEFNI-PIVTGWWGCGA   | 683              |                  |             |
| QY | 867                   | FGDARLKAIIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVGKVKLLLR    | 926              |                  |             |
| DB | 684                   | FNGDKPLKFIQIVIAAGVADRPLHFCFSFGEPELAACKKIIIERMKQKDVTLGMLFSMI-- | 741              |                  |             |
| QY | 927                   | YNYEECRNCSTPGPDIKLYPF   | 947              |                  |             |
| DB | 742                   | -----NNTGLPHKHFEFYVE  | 756              |                  |             |

|                |                                       |
|----------------|---------------------------------------|
| RESULT 15      |                                       |
| Q19637         |                                       |
| ID Q19637      | PRELIMINARY; PRT; 764 AA.             |
| AC Q19637;     |                                       |
| DT 01-NOV-1996 | (TREMBLrel. 01, Created)              |
| DT 01-MAR-2003 | (TREMBLrel. 23; Last sequence update) |





GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 77.8912 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-4  
Perfect score: 5190  
Sequence: 1 MNAGPGCEPCTKATRWGAAT.....YHAVESCAETADHSGQRTGT 976

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID     | Description         |
|------------|--------|-------------|--------|-----------|---------------------|
| 1          | 5190   | 100.0       | 976    | 4 Q9Y4W7  | Q9Y4W7 homo sapien  |
| 2          | 5113   | 98.5        | 976    | 4 Q86W56  | Q86W56 homo sapien  |
| 3          | 5108   | 98.4        | 976    | 4 Q7Z742  | Q7Z742 homo sapien  |
| 4          | 4635.5 | 89.3        | 977    | 6 O02776  | O02776 bos taurus   |
| 5          | 4395   | 84.7        | 968    | 11 O88622 | O88622 mus musculus |
| 6          | 4381.5 | 84.4        | 961    | 11 Q80YQ6 | Q80YQ6 mus musculus |
| 7          | 4336   | 83.5        | 972    | 11 Q9QYM2 | Q9QYM2 rattus norv  |
| 8          | 4165.5 | 80.3        | 920    | 11 Q8CB72 | Q8CB72 mus musculus |
| 9          | 1078   | 20.8        | 768    | 5 O46043  | O46043 drosophila   |
| 10         | 1073   | 20.7        | 723    | 5 Q960N8  | Q960N8 drosophila   |
| 11         | 579    | 11.2        | 548    | 10 Q9SKB3 | Q9SKB3 arabidopsis  |
| 12         | 539    | 10.4        | 522    | 10 Q8VYA1 | Q8VYA1 arabidopsis  |
| 13         | 480.5  | 9.3         | 781    | 5 Q867X0  | Q867X0 caenorhabdi  |
| 14         | 473    | 9.1         | 764    | 5 Q19637  | Q19637 caenorhabdi  |
| 15         | 453    | 8.7         | 485    | 5 Q9N5L4  | Q9N5L4 caenorhabdi  |
| 16         | 340    | 6.6         | 364    | 10 Q9SKB4 | Q9SKB4 arabidopsis  |

|    |       |     |      |           |                    |
|----|-------|-----|------|-----------|--------------------|
| 17 | 333.5 | 6.4 | 368  | 5 Q86GI4  | Q86GI4 toxoplasma  |
| 18 | 188   | 3.6 | 508  | 12 Q9E234 | Q9E234 helicoverpa |
| 19 | 184   | 3.5 | 510  | 12 Q99GU9 | Q99GU9 helicoverpa |
| 20 | 178   | 3.4 | 2081 | 10 Q9LH98 | Q9LH98 arabidopsis |
| 21 | 175   | 3.4 | 1078 | 5 Q963T1  | Q963T1 plasmodium  |
| 22 | 172   | 3.3 | 2241 | 5 Q8IK37  | Q8IK37 plasmodium  |
| 23 | 171   | 3.3 | 744  | 5 Q86S96  | Q86S96 babesia rod |
| 24 | 170.5 | 3.3 | 4969 | 11 Q8CF91 | Q8CF91 mus musculu |
| 25 | 170.5 | 3.3 | 5165 | 11 Q8CF92 | Q8CF92 mus musculu |
| 26 | 170   | 3.3 | 5322 | 5 Q9VPL9  | Q9VPL9 drosophila  |
| 27 | 169.5 | 3.3 | 1274 | 16 Q8NW31 | Q8NW31 staphylococ |
| 28 | 169   | 3.3 | 498  | 5 Q8MTN8  | Q8MTN8 trichinella |
| 29 | 168   | 3.2 | 3167 | 5 Q17464  | Q17464 caenorhabdi |
| 30 | 167.5 | 3.2 | 968  | 5 Q8I5W9  | Q8I5W9 plasmodium  |
| 31 | 167   | 3.2 | 1988 | 5 Q9VU16  | Q9VU16 drosophila  |
| 32 | 166.5 | 3.2 | 1163 | 4 Q9UHB7  | Q9UHB7 homo sapien |
| 33 | 166.5 | 3.2 | 1472 | 5 Q8I2A4  | Q8I2A4 plasmodium  |
| 34 | 165.5 | 3.2 | 571  | 5 Q8MTN7  | Q8MTN7 trichinella |
| 35 | 165.5 | 3.2 | 1274 | 16 Q99TC3 | Q99TC3 staphylococ |
| 36 | 165.5 | 3.2 | 1444 | 5 Q9VTN2  | Q9VTN2 drosophila  |
| 37 | 165.5 | 3.2 | 1514 | 5 Q8SY55  | Q8SY55 drosophila  |
| 38 | 165   | 3.2 | 838  | 10 Q9CAK9 | Q9CAK9 arabidopsis |
| 39 | 165   | 3.2 | 884  | 10 Q9CAB4 | Q9CAB4 arabidopsis |
| 40 | 164   | 3.2 | 4717 | 3 Q94248  | Q94248 schizosacch |
| 41 | 163.5 | 3.2 | 864  | 16 Q7ZAM9 | Q7ZAM9 leptospira  |
| 42 | 163.5 | 3.2 | 1182 | 5 Q8ID30  | Q8ID30 plasmodium  |
| 43 | 163   | 3.1 | 349  | 6 Q8HY98  | Q8HY98 centurio se |
| 44 | 162   | 3.1 | 455  | 5 Q86A18  | Q86A18 dictyosteli |
| 45 | 162   | 3.1 | 585  | 5 Q8IG74  | Q8IG74 caenorhabdi |

ALIGNMENTS

RESULT 1

|        |   |              |      |         |
|--------|---|--------------|------|---------|
| Q9Y4W7 | Q9Y4W7  | PRELIMINARY; | PRT; | 976 AA. |
| ID     | Q9Y4W7  |              |      |         |
| AC     | Q9Y4W7;   |              |      |         |
| DT     | 01-NOV-1999 (TREMBlrel. 12, Created)  |              |      |         |
| DT     | 01-NOV-1999 (TREMBlrel. 12, Last sequence update)   |              |      |         |
| DT     | 01-JUN-2003 (TREMBlrel. 24, Last annotation update)   |              |      |         |
| DE     | Poly(ADP-ribose) glycohydrolase.  |              |      |         |
| GN     | HPARG.  |              |      |         |
| OS     | Homo sapiens (Human).   |              |      |         |
| OC     | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |              |      |         |
| OC     | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |              |      |         |
| OX     | NCBI_TaxID=9606;  |              |      |         |
| RN     | [1]   |              |      |         |
| RP     | SEQUENCE FROM N.A.  |              |      |         |
| RX     | MEDLINE=9380098; PubMed=10449915;   |              |      |         |
| RA     | Ame J.C., Apicu F., Jacobson E.L., Jacobson M.K.;   |              |      |         |
| RT     | "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to human chromosome 10q11.23 and mouse chromosome 14B by in situ hybridization." |              |      |         |
| RL     | Cytogenet. Cell Genet. 85:269-270 (1999).   |              |      |         |
| DR     | EMBL; AF005043; AAB61614.1; -.  |              |      |         |
| DR     | Genew; HGNC:8605; PARG.   |              |      |         |
| DR     | GO; GO:0005737; C:cytoplasm; TAS.   |              |      |         |
| DR     | GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  |              |      |         |
| DR     | InterPro; IPR007724; PARG.  |              |      |         |
| DR     | Pfam; PF05028; PARG; 1.   |              |      |         |
| KW     | Hydrolase.  |              |      |         |
| SQ     | SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;   |              |      |         |

Query Match 100.0%; Score 5190; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |    |   |     |
|----|----|---|-----|
| QY | 1  | MNAGPGCEPCTKATRWGAATTSPAASDARSFPPSRQRRVLDPKDAHVQFRVPPSSPACVPG | 60  |
| Db | 1  | MNAGPGCEPCTKATRWGAATTSPAASDARSFPPSRQRRVLDPKDAHVQFRVPPSSPACVPG | 60  |
| QY | 61 | QAGQHRGSATSLVFKQKTTTSMWDTKGIKTAESESLDSKENNTRIESMMSSVQKDNFYQ   | 120 |

|    |     |  |     |
|----|-----|--|-----|
| Db | 61  | QAGQHRGSATSLVPKQKTTITSWMDTKIGIKTAESSELDSEKNNNTRIESMMSSVQKDNFYQ | 120 |
| QY | 121 | HNVEKLVNVSQSLDKSLTEKSTQYLNQHQTAAAMCKWQNEGKHTEQLLESEPQTVLVPE    | 180 |
| Db | 121 | HNVEKLVNVSQSLDKSLTEKSTQYLNQHQTAAAMCKWQNEGKHTEQLLESEPQTVLVPE    | 180 |
| QY | 181 | QFSNANIDRSPQNDHSDTDSEENRDNQOFLTTVKLANAKQTTTEDEHAREAKSHQKCSKS   | 240 |
| Db | 181 | QFSNANIDRSPQNDHSDTDSEENRDNQOFLTTVKLANAKQTTTEDEHAREAKSHQKCSKS   | 240 |
| QY | 241 | CHPGEDCASCQQDEIDVVPKSPKSDVGSSEVGTGSKNDNKLIRQESCLGNSPPFEKESEP   | 300 |
| Db | 241 | CHPGEDCASCQQDEIDVVPKSPKSDVGSSEVGTGSKNDNKLIRQESCLGNSPPFEKESEP   | 300 |
| QY | 301 | ESPMDDVNSKNSCQDSEADETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK    | 360 |
| Db | 301 | ESPMDDVNSKNSCQDSEADETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK    | 360 |
| QY | 361 | GGEVRLHFQFEGGESRTGMNDLNAKLPGNISLNVCECRNSKQHGKDKSKITDHLMLPKA    | 420 |
| Db | 361 | GGEVRLHFQFEGGESRTGMNDLNAKLPGNISLNVCECRNSKQHGKDKSKITDHLMLPKA    | 420 |
| QY | 421 | EDRRKEQWETKHQRTERKIPKYVPPHLSPKDKWLGTPIEEMRMRPCGIRLPLLRPSANH    | 480 |
| Db | 421 | EDRRKEQWETKHQRTERKIPKYVPPHLSPKDKWLGTPIEEMRMRPCGIRLPLLRPSANH    | 480 |
| QY | 481 | TVTIRVDLLRAGEVPKPPFPTHYKDLWDNKHVKMPCSEONLYPVEDENGERTAGSRWELIQ  | 540 |
| Db | 481 | TVTIRVDLLRAGEVPKPPFPTHYKDLWDNKHVKMPCSEONLYPVEDENGERTAGSRWELIQ  | 540 |
| QY | 541 | TALLNKFTRPQNLKDAILKYNVAYSKKWDF TALIDFWDKVL EEAQAHL YQSILPDMVKI | 600 |
| Db | 541 | TALLNKFTRPQNLKDAILKYNVAYSKKWDF TALIDFWDKVL EEAQAHL YQSILPDMVKI | 600 |
| QY | 601 | ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYPDIN   | 660 |
| Db | 601 | ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYPDIN   | 660 |
| QY | 661 | FNRLFEGRSSRKPEKLKTLFCYFRFVTEKKPTGLVFTTQOSLEDFFPEWERCEKPLTRLHV  | 720 |
| Db | 661 | FNRLFEGRSSRKPEKLKTLFCYFRFVTEKKPTGLVFTTQOSLEDFFPEWERCEKPLTRLHV  | 720 |
| QY | 721 | TYEGTIEENGQGMLOQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNEC  | 780 |
| Db | 721 | TYEGTIEENGQGMLOQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNEC  | 780 |
| QY | 781 | LIITGTEQYSEYTGAEITYRWSRSHEDGSEDDCERRCTEIVAIDALHFRRYLDQFVPEK    | 840 |
| Db | 781 | LIITGTEQYSEYTGAEITYRWSRSHEDGSEDDCERRCTEIVAIDALHFRRYLDQFVPEK    | 840 |
| QY | 841 | MRRELKAYCGFLRPGVSSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAAERDWWYF   | 900 |
| Db | 841 | MRRELKAYCGFLRPGVSSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAAERDWWYF   | 900 |
| QY | 901 | TFGDSELMRDIYSMHIFLTERKLTVGDDVYKLLRLRYNNEECRNCSTPGPDIKLYPFYHAV  | 960 |
| Db | 901 | TFGDSELMRDIYSMHIFLTERKLTVGDDVYKLLRLRYNNEECRNCSTPGPDIKLYPFYHAV  | 960 |
| QY | 961 | ESCAETADHSGORTGT   | 976 |
| Db | 961 | ESCAETADHSGORTGT   | 976 |

## RESULT 2

|                |                              |                         |         |
|----------------|------------------------------|-------------------------|---------|
| Q86W56         | PRELIMINARY;                 | PRT;                    | 976 AA. |
| ID Q86W56      |                              |                         |         |
| AC Q86W56;     |                              |                         |         |
| DT 01-JUN-2003 | (TREMBLrel. 24,              | Created)                |         |
| DT 01-JUN-2003 | (TREMBLrel. 24,              | Last sequence update)   |         |
| DT 01-OCT-2003 | (TREMBLrel. 25,              | Last annotation update) |         |
| DE             | Similar to poly (ADP-ribose) | glycohydrolase.         |         |
| OS             | Homo sapiens (Human).        |                         |         |

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP  
RC TISSUE=Skin;  
RA Straussberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; BC050560; AAH50560.1; --  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
  
Query Match 98.5%; Score 5113; DB 4; Length 976;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 963: Conservative 3: Mismatches 10: Indels 0: Gaps 0;

2

Db 781 LIITGEQYSEYTGAEYRWSRSHEDGSRDDWQRCTEIVAIDALHFRRYLDQFVPEK 840

QY 841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYF 900

Db 841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYF 900

QY 901 TFGDSELMRDIYSMHIFLTERKLTVDVYKLLRYYNEECRCNSTPGPDIKLYPFIYHAV 960

Db 901 TFGDSELMRDIYSMHIFLTERKLTVDVYKLLRYYNEECRCNSTPGPDIKLYPFIYHAV 960

QY 961 ESCAETADHSGQRTGT 976

Db 961 ESCAETADHSGQRTGT 976

RESULT 3

Q7Z742 PRELIMINARY; PRT; 976 AA.

ID Q7Z742

AC Q7Z742;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA Strausberg R.;

DR Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

KW EMBL; BC052966; AAH52966.1; -.

KW Hypothetical protein.

SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;

Query Match 98.4%; Score 5108; DB 4; Length 976;

Best local Similarity 98.6%; Pred. No. 0;

Matches 962; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MNAGPGCEPCTKATFWGAATTSPAASDARSFRRVLDPKDAHVFRRVPPSSPACVPG 60

Db 1 MNAGPGCEPCTKATFWGAATTSPAASDARSFRRVLDPKDAHVFRRVPPSSPACVPG 60

QY 61 QAGQHRGATSLSVFKQKTIISWMDTKGIKTAESLDSKENNTRIESMMSSVQKDNFYQ 120

Db 61 RAGQHRGATSLSVFKQKTIISWMDTKGIKTAESLDSKENNTRIESMMSSVQKDNFYQ 120

QY 121 HNVEKLVNSQLSLDKSTEKSTQYLNQHTAAMCKWQNEGKHTEQLESEPTVTLVPE 180

Db 121 HNVEKLVNSQLSLDKSTEKSTQYLNQHTAAMCKWQNEGKHTEQLESEPTVTLVPE 180

QY 181 QFSNANIDRSPQNDHSDTDSEENRDNQFLTTTVKLANAKQTTEDEHAREAKSHQCKSKS 240

Db 181 QFSNANIDRSPQNDHSDTDSEENRDNQFLTTTVKLANAKQTTEDEHAREAKSHQCKSKS 240

QY 241 CHPGEDCASCQQDEIDVVPKSPSLSDVGSDEVGTGSKNDKNKLIRQESCLGNSPPFEKESEP 300

Db 241 CDPGEDCASCQQDEIDVVPKSPSLSDVGSDEVGTGSKNDKNKLIRQESCLGNSPPFEKESEP 300

QY 301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQANKPSRFOARDADIEFRKRYSTK 360

Db 301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQANKPSRFOARDADIEFRKRYSTK 360

QY 361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISLNVCECRNSKHGKKDSKITDHLMLRPKA 420

Db 361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISLNVCECRNSKHGKKDSKITDHLMLRPKA 420

QY 421 EDRRKEQWETKHQRTERKIPKYVPPHLSPPDKKWLGTPIEMRRMRPCGIRLPLLRPSANH 480

Db 421 EDRRKEQWETKHQRTERKIPKYVPPHLSPPDKKWLGTPIEMRRMRPCGIRLPLLRPSANH 480

QY 481 TVTIRVDLLRAGEVPKPPFPHYKDLWDNKHVMPCSEONLYPVEDENGERTAGSRWELIQ 540

Db 481 TVTIRVDLLRAGEVPKPPFPHYKDLWDNKHVMPCSEONLYPVEDENGERTAGSRWELIQ 540

QY 541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLBEAEQAHLQYQILPDMVKI 600

Db 541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLBEAEQAHLQYQILPDMVKI 600

QY 601 ALCLPNICTOPIPLLKQKWNHITMSQEQIASLLANAFCTFPRRNMKMKSEYSSYPDIN 660

Db 601 ALCLPNICTOPIPLLKQKWNHITMSQEQIASLLANAFCTFPRRNMKMKSEYSSYPDIN 660

QY 661 FNRLEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDFFPEWERCEKPLTRLHV 720

Db 661 FNRLEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDFFPEWERCEKPLTRLHV 720

QY 721 TYEGTIEENGQGMQLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNEC 780

Db 721 TYEGTIEENGQGMQLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNEC 780

QY 781 LIITGEQYSEYTGAEYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Db 781 LIITGEQYSEYTGAEYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

QY 841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYF 900

Db 841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYF 900

QY 901 TFGDSELMRDIYSMHIFLTERKLTVDVYKLLRYYNEECRCNSTPGPDIKLYPFIYHAV 960

Db 901 TFGDSELMRDIYSMHIFLTERKLTVDVYKLLRYYNEECRCNSTPGPDIKLYPFIYHAV 960

QY 961 ESCAETADHSGQRTGT 976

Db 961 ESCAETADHSGQRTGT 976

RESULT 4

O02776 PRELIMINARY; PRT; 977 AA.

ID O02776

AC O02776;

DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Poly(ADP-ribose) glycohydrolase.

GN BPARG.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;

OC Bovidae; Bovinae; Bos.



```

OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97277328; PubMed=9115250;
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding bovine poly (ADP-
RT ribose) glycohydrolase."
RL J. Biol. Chem. 272:11895-11901 (1997).
DR EMBL; U78975; AAB53370.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;

Query Match 89.3%; Score 4635.5; DB 6; Length 977;
Best Local Similarity 89.5%; Pred. No. 0;
Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;

QY 1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPPSRQRRVLDPKDAHVFQFRVPPSSPACVP 59
DB 1 MSAGPGCEPCTKPRWDAAATSPPAASDARSFPPGRQRRVLDSPKAPVQFRVPPSSSGCAL 60

QY 60 GOAGQHRGSATSLVFKQKTTITSMWDTKGIKTAESLSDSKENNNTRIESMSSSVQKDNFY 119
DB 61 GRAGQHRGSATSLVFKQKTTITSMWDTKGIKTAESLSDSKENNNTRIESMSSSVQKDNFY 120

QY 120 QHNVEKLVNVSQSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTTLVP 179
DB 121 QHNMEKLVNVSQSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTTLVP 180

QY 180 EQFSNANIDRSPQNDHSDTDSEENRDNQOFLTIVKLANAKQTTEDHAREAKSHQKCSK 239
DB 181 EQFSNANVDQSPKDDHSDTSEESRDNQOFLTIVKLANAKQTTEDHAREAKSHQKCGK 240

QY 240 SCHPGEDCASCQDEIDVWPKSPSLSDVGSSEDTGSKNDNKLIRQESCLGNSPPFEKES 299
DB 241 ACHPAEACAGCQDETDVVSSEPLSDTGSSEDTGSKNANRLNRQESSLGNSPPFEKES 300

QY 300 PESPMVDVNSKNSCODSEADEETSPGFDEQEDSSSQATANKPSRFQARDADIEFRKRYST 359
DB 301 PESPMVDVNSKNSCODSEADEETSPGFDEQEDSSSQATANKPSRFQARDADIEFRKRYST 360

QY 360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKDKSKITDHLMLRPK 419
DB 361 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKDKSKITDHLMLRPK 420

QY 420 AEDRRKEQWETKHQRTKIPKYVPPHLSQKWLGTPIEMRRMPCGIRLPLRPSAN 479
DB 421 AEDKRKEQCEMKHQRTKIPKYVPPHLSQKWLGTPIEMRRMPCGIRLPLRPSAN 480

QY 480 HTVTIRVDLLRAGEVPPKPTTHYKDLNDKHKVMPCEQNLVYVDENGERTAGSRWELI 539
DB 481 HTVTIRVDLLRAGEVPPKPTTHYKDLNDKHKVMPCEQNLVYVDENGERTAGSRWELI 540

QY 540 QTALLNKFTRPQNLKDAILLKYNVAYSKKWDF TALIDFWDKVL EEAQAHLVQSILPDMVK 599
DB 541 QTALLNRLTRPQNLKDAILLKYNVAYSKKWDF TALIDFWDKVL EEAQAHLVQSILPDMVK 600

QY 600 IALCLPNICTQPIPLLKQKNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659
DB 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660

QY 660 NFNRLFEGRSRKPPEKLTLCYFRRVTEKKPTGLVTFTRQSLDFPEWERCXPLTRLH 719
DB 661 NFNRLFEGRSRKPPEKLTLCYFRRVTEKKPTGLVTFTRQSLDFPEWERCXPLTRLH 720

QY 720 VTYEGTIEENGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLHDNE 779
DB 721 VTYEGTIEENGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLHDNE 780

QY 780 CLITGTEQYSEYTGAEYRWARSHEDRSERDDQWRRTTEIVAIDALHFERRYLDFQVPE 840
DB 781 CLITGTEQYSEYTGAEYRWARSHEDRSERDDQWRRTTEIVAIDALHFERRYLDFQVPE 840

QY 840 KMRRELNKAYCGFLRPGVSSSENLSAVATGNWCGAFGSDARLKALIQILAAAAAERDVVY 899
DB 841 KIRRELNKAYCGFLRPGVSSSENLSAVATGNWCGAFGSDARLKALIQILAAAAAERDVVY 900

QY 900 FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHA 959
DB 901 FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHA 960

QY 960 VESCAETADHSGQRTG 975
DB 961 VESCTQTTNQPGQRTG 976

RESULT 5
O88622 PRELIMINARY; PRT; 968 AA.
AC O88622;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380098; PubMed=10449915;
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT hybridization."
RL Cytogenet. Cell Genet. 85:269-270 (1999).
DR EMBL; AF079557; AAC28735.1; -.
DR MGD; MGI:1347094; Parg.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;

Query Match 84.7%; Score 4395; DB 11; Length 968;
Best Local Similarity 85.7%; Pred. No. 1.1e-306;
Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;

QY 1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPPSRQRRVLDPKDAHVFQFRVPPSSPACVP 59
DB 1 MSAGPGWEPTKA-RWGAAGTSAPTASDSRSFPPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

QY 60 GOAGQHRGSATSLVFKQKTTITSMWDTKGIKTAESLSDSKENNNTRIESMSSSVQKDNFY 119
DB 60 GRAGPHRGNAISFVFKQKTTITWMDTKGPKTAESE---SKENNNTRIDSMSSSVQKDNFY 116

QY 120 QHNVEKLVNVSQSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTTLVP 179
DB 117 PHKVEKLVNVSQSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTTLVP 176

QY 180 EQFSNANIDRSPQNDHSDTDSEENRDNQOFLTIVKLANAKQTTEDHAREAKSHQKCSK 239
DB 177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQOFLTIVKLANAKQTTEDHAREAKSHQKCSK 233

QY 240 SCHPGEDCASCQDEIDVWPKSPSLSDVGSSEDTGSKNDNKLIRQESCLGNSPPFEKES 299
DB 234 SRQSVKDCGCGQEEVDVLPESPLSDVGAEDIGTGPKNDNKLITGQESSLGDSPPFEKES 293

QY 300 PESPMVDVNSKNSCODSEADEETSPGFDEQEDSSSQATANKPSRFQARDADIEFRKRYST 359
DB 294 PESPMVDVNSKNSCODSEADEETSPGFDEQEDSSSQATANKPSRFQARDADIEFRKRYST 352

QY 360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKDKSKITDHLMLRPK 419
```

```

DB 781 CLITGTEQYSEYTGAEYRWARSHEDRSERDDQWRRTTEIVAIDALHFERRYLDFQVPE 840
QY 840 KMRRELNKAYCGFLRPGVSSSENLSAVATGNWCGAFGSDARLKALIQILAAAAAERDVVY 899
DB 841 KIRRELNKAYCGFLRPGVSSSENLSAVATGNWCGAFGSDARLKALIQILAAAAAERDVVY 900
QY 900 FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHA 959
DB 901 FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHA 960
QY 960 VESCAETADHSGQRTG 975
DB 961 VESCTQTTNQPGQRTG 976

RESULT 5
O88622 PRELIMINARY; PRT; 968 AA.
AC O88622;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380098; PubMed=10449915;
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT hybridization."
RL Cytogenet. Cell Genet. 85:269-270 (1999).
DR EMBL; AF079557; AAC28735.1; -.
DR MGD; MGI:1347094; Parg.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;

Query Match 84.7%; Score 4395; DB 11; Length 968;
Best Local Similarity 85.7%; Pred. No. 1.1e-306;
Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;

QY 1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPPSRQRRVLDPKDAHVFQFRVPPSSPACVP 59
DB 1 MSAGPGWEPTKA-RWGAAGTSAPTASDSRSFPPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

QY 60 GOAGQHRGSATSLVFKQKTTITSMWDTKGIKTAESLSDSKENNNTRIESMSSSVQKDNFY 119
DB 60 GRAGPHRGNAISFVFKQKTTITWMDTKGPKTAESE---SKENNNTRIDSMSSSVQKDNFY 116

QY 120 QHNVEKLVNVSQSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTTLVP 179
DB 117 PHKVEKLVNVSQSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTTLVP 176

QY 180 EQFSNANIDRSPQNDHSDTDSEENRDNQOFLTIVKLANAKQTTEDHAREAKSHQKCSK 239
DB 177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQOFLTIVKLANAKQTTEDHAREAKSHQKCSK 233

QY 240 SCHPGEDCASCQDEIDVWPKSPSLSDVGSSEDTGSKNDNKLIRQESCLGNSPPFEKES 299
DB 234 SRQSVKDCGCGQEEVDVLPESPLSDVGAEDIGTGPKNDNKLITGQESSLGDSPPFEKES 293

QY 300 PESPMVDVNSKNSCODSEADEETSPGFDEQEDSSSQATANKPSRFQARDADIEFRKRYST 359
DB 294 PESPMVDVNSKNSCODSEADEETSPGFDEQEDSSSQATANKPSRFQARDADIEFRKRYST 352

QY 360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKDKSKITDHLMLRPK 419
```

|    |     |   |     |
|----|-----|---|-----|
| Db | 353 | KGSEVRLHFQFE - GENNAGTSDLNAKPSGNSSLNVECRSSKQHGKRD SKITDHFMRISK  | 411 |
| QY | 420 | AEDRRKEQWETKHQRTKIPKYVPPHLSDPDKWLGTPIEEMRRMPCRGIRLPLRPSAN       | 479 |
| Db | 412 | SEDRRKEQCEVRHQRTKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS      | 471 |
| QY | 480 | HTVTIRVDLLRAGEVPKPPPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI    | 539 |
| Db | 472 | HTVTIRVDLLRAGEVPKPPPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI    | 531 |
| QY | 540 | QTALLNKFTRPQNLKDAILKYNVAYSKKWDF TALIDFWDKVL EEAQAHL YQSILPDMVK  | 599 |
| Db | 532 | QTALLNKFTRPQNLKDAILKYNVAYSKKWDF TALIDFWDKVL EEAQAHL YQSILPDMVK  | 591 |
| QY | 600 | IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNNAKMKSEYSSYPDI    | 659 |
| Db | 592 | IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFCTFPRRNNAKMKSEYSSYPDI    | 651 |
| QY | 660 | NFNRLFEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFRQSLEDFPEWERCCKPLTRLH      | 719 |
| Db | 652 | NFNRLFEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFRQSLEDFPEWERCCKPLTRLH      | 711 |
| QY | 720 | VTYEGTIEENGQGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNE    | 779 |
| Db | 712 | VTYEGTIEENGQGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNE    | 771 |
| QY | 780 | CLIIITGTEQYSEYTGAEYRWRSRSHEDGSEKDDCERRCTEIVAIDALHFRRYLDQFVPE    | 839 |
| Db | 772 | CLIIITGTEQYSEYTGAEYRWRSRSHEDGSEKDDCERRCTEIVAIDALHFRRYLDQFVPE    | 831 |
| QY | 840 | KMRRELNKAYCGFLRPGVPSSENL SAVATGNWGC GAFGGDARLKALIQILAAAAAERDVVY | 899 |
| Db | 832 | KVRRELNKAYCGFLRPGVPSSENL SAVATGNWGC GAFGGDARLKALIQILAAAAAERDVVY | 891 |
| QY | 900 | FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHA   | 959 |
| Db | 892 | FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHA   | 951 |
| QY | 960 | VESCAETADHSGQRTGT 976   |     |
| Db | 952 | VESSAETDMPGQKAGT 968  |     |

RESULT 6  
Q80YQ6 PRELIMINARY; PRT; 961 AA.  
AC Q80YQ6;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050892; AAHS0892.1; -.  
DR GO; GO:0016787; F.hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

Query Match 84.4%; Score 4381.5; DB 11; Length 961;  
Best Local Similarity 86.0%; Pred. No. 9.7e-306;  
Matches 833; Conservative 50; Mismatches 77; Indels 9; Gaps 5;  
QY 1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

|    |     |   |     |
|----|-----|---|-----|
| Db | 1   | MSAGPGWEPCTKRPRWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPSSPACVS     | 60  |
| QY | 60  | GOAGQHRGATSILVFKQKTITSWMDTKIGKTAESLSDSKENNNTRIESMMSSVQKDNFY     | 119 |
| Db | 61  | GRAGPHRGNATSIVFKQKTITWMDTKGPKTAESE--SKENNNTRIDSMSSVQKDNFY       | 117 |
| QY | 120 | QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHTAAMCKWQNEGKHTEQLLESEPTVTLPV      | 179 |
| Db | 118 | PHKVEKLENVPLNLDKSPTEKSSQYLNQQTASVCKWQNEGKHAEQLLASEPPAGTPLP      | 177 |
| QY | 180 | EQFSNANIDRSPQNDHSDTDSEENRDNQOFLTTVVKLANAKQTTTEDEHAREAKSHQKCSK   | 239 |
| Db | 178 | KQLSNANIGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDGQAR---SNCKCSG    | 234 |
| QY | 240 | SCHPGEDCASCQQDEIDVVPKSPLSVGSSEVGTGSKNDNKLIRQESCLGNSPPFEKESE     | 299 |
| Db | 235 | SRQSVKDCCTGCQEEVDVLPESPLSDVGAEDIGTGPKNNDNKLGTQESSLGDSPFEKESE    | 294 |
| QY | 300 | PESPMDVDNSKNSCODSEADEETSPGDFEQEDGSSSQTANKPSRFQARDADIERKRYST     | 359 |
| Db | 295 | PESPMDVDNSKNSCODSEADEETSPVFEQDD-RSSQTANKLSSCQAREADGDLRKRYLT     | 353 |
| QY | 360 | KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKDKSKITDHLMLPK     | 419 |
| Db | 354 | KGSEVRLHFQFE-GENNVGTSDLNAKPSGNSSSLNVECRSSKQHGKDKSKITDHFMRISK    | 412 |
| QY | 420 | AEDRRKEQWETKHQRTKIPKYVPPHLSDPDKWLGTPIEEMRRMPCRGIRLPLRPSAN       | 479 |
| Db | 413 | SEDRRKEQCEVRHQRTKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS      | 472 |
| QY | 480 | HTVTIRVDLLRAGEVPKPPPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI    | 539 |
| Db | 473 | HTVTIRVDLLRAGEVPKPPPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI    | 532 |
| QY | 540 | QTALLNKFTRPQNLKDAILKYNVAYSKKWDF TALIDFWDKVL EEAQAHL YQSILPDMVK  | 599 |
| Db | 533 | QTALLNKFTRPQNLKDAILKYNVAYSKKWDF TALIDFWDKVL EEAQAHL YQSILPDMVK  | 592 |
| QY | 600 | IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNNAKMKSEYSSYPDI    | 659 |
| Db | 593 | IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFCTFPRRNNAKMKSEYSSYPDI    | 652 |
| QY | 660 | NFNRLFEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFRQSLEDFPEWERCCKPLTRLH      | 719 |
| Db | 653 | NFNRLFEGRSSRKPEKLTLCYFRRRVTEKKPTGLVTFRQSLEDFPEWERCCKPLTRLH      | 712 |
| QY | 720 | VTYEGTIEENGQGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNE    | 779 |
| Db | 713 | VTYEGTIEGNRGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDHNE     | 772 |
| QY | 780 | CLIIITGTEQYSEYTGAEYRWRSRSHEDGSEKDDCERRCTEIVAIDALHFRRYLDQFVPE    | 839 |
| Db | 773 | CLIIITGTEQYSEYTGAEYRWRSRSHEDGSEKDDCERRCTEIVAIDALHFRRYLDQFVPE    | 832 |
| QY | 840 | KMRRELNKAYCGFLRPGVPSSENL SAVATGNWGC GAFGGDARLKALIQILAAAAAERDVVY | 899 |
| Db | 833 | KVRRELNKAYCGFLRPGVPSSENL SAVATGNWGC GAFGGDARLKALIQILAAAAAERDVVY | 892 |
| QY | 900 | FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHA   | 959 |
| Db | 893 | FTFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHA   | 952 |
| QY | 960 | VESCAETAD 968   |     |
| Db | 953 | VESSAETTD 961   |     |

RESULT 7  
Q9QYM2 PRELIMINARY; PRT; 972 AA.  
ID Q9QYM2  
AC Q9QYM2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)





Db 235 SRQSVKDTGCGQEEVDVLPESPLSDVGAEDIGTGPKNDNKLITGOESSLGDSPPEKESE 294

QY 300 PESPMVDVNSKNSQDSEADEETSPGFDEQEDGSSQTANKPSRFQARDADIEFRKRYST 359

Db 295 PESPMVDVNSKNSQDSEADEETSPVFDEQDD-RSQTANKLSSCQAREADGDLKRYLT 353

QY 360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISLNVCECRNSKHQKDKSKITDHLMLPK 419

Db 354 KGSEVRLHFQFE-GENNAGTSDLNAPSGNSSSLNVECRSSKHQKDRSKITDHFMRISK 412

QY 420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPOKWLGTPIEEMRMRPCGIRLPLRPSAN 479

Db 413 SEDRRKEQCEVRHQRTERKIPKYVPPHLSPOKWLGTPIEEMRMRPCGIRLPLRPSAS 472

QY 480 HTVTIRVDLLRAGEVPKPFPTHYKDLWMDNKHVMKPCSEONLYPVEDENGERTAGSRWELI 539

Db 473 HTVTIRVDLLRAGEVPKPFPTHYKDLWMDNKHVMKPCSEONLYPVEDENGERTAGSRWELI 532

QY 540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLSEAEAOHLYQSILPDMVK 599

Db 533 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLSEAEAOHLYQSILPDMVK 592

QY 600 IALCLPNICTQPIPLLKQNMHSITMSQEQIASLLANAFCTFPPRNAKMKSEYSSYPDI 659

Db 593 IALCLPNICTQPIPLLKQNMHSITMSQEQIASLLANAFCTFPPRNAKMKSEYSSYPDI 652

QY 660 NFNRLFEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDFFPEWERCEKPLTRLH 719

Db 653 NFNRLFEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDFFPEWERCEKPLTRLH 712

QY 720 VTYEGTIEENGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISLFTVEVDHNE 779

Db 713 VTYEGTIEENGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISLFTVEVDHNE 772

QY 780 CLITGTQYSEYTYAETRWSPSHEDGSDERDDCERRCTEIVADALHFRYLDQFVPE 839

Db 773 CLITGTQYSEYTYAETRWSPSHEDGSDERDDCERRCTEIVADALHFRYLDQFVPE 832

QY 840 KMRRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVY 899

Db 833 KVRRELKAYCGFLRPGVPSNLSAVATGNWCGAFGGDVRKALIQILAAAAAERDVVY 892

QY 900 FTFGSELMDIYSMHIFLTERKLTIVGD 927

Db 893 FTFGSELMDIYSMHIFLTERKLTIVGD 920

RESULT 9

O46043

ID O46043 PRELIMINARY; PRT; 768 AA.

AC O46043;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE PARG protein.

GN PARG OR EG:114E2.1 OR CG2864.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Ame J.-C., Jacobson M.K.;

RT "Isolation and characterization of the cDNA encoding Drosophila

RL poly(ADP-ribose) glycohydrolase.";

RN Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RP [3]

RP SEQUENCE OF 46-768 FROM N.A.

RA Murphy L., Harris D., Barrell B.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 46-768 FROM N.A.

RA Benos P.;

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003428; AAF45886.1; -

DR EMBL; AF079556; AAC28734.1; -

DR EMBL; Z98254; CAB10913.1; -

DR FlyBase; FBgn0023216; Parg.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

KW Hydrolase.

SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;

Query Match 20.8%; Score 1078; DB 5; Length 768;

Best Local Similarity 42.3%; Pred. No. 1.3e-68;

Matches 240; Conservative 94; Mismatches 184; Indels 50; Gaps 14;

QY 416 RLPKAEARRRKEQWETKHQRTERKIPKYVPPHL--SPDKKWLGTPIEEM--RRMPCGIRL 471

Db 45 RMSKSPDGGISEIETEE-----PENLANSLDSDNRGVSMEAIHNRNQPFELENL 94

QY 472 PLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWMDNKHVMKPCSEONLYPVEDENGERT 531

Db 95 PPVTAGNLHRVMYQLPIRET--PPRPYKSPGK--WDSEHVLPLCAPESKYPRENPDGSTT 150

QY 532 AGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLSEAEAOHLYQ 591

Db 151 IDFRWEMIERALLQPIKTCELOAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

QY 592 SILPDMVKIALCLPNICTQPIPLLKQNMHSITMSQEQIASLLANAFCTFPPRNA-KMK 650

Db 211 DLLPRIIRIALRLPDLIQSPVPLLKHKHKNASLSLSQQQISCLLANAFCTFPPRNTLKRK 270





Db 139 VFLSQELIGALLACSFCLFPDDNRGAK-----HLPVINFDHLFASLIYSYSQSQESKIRC 194  
QY 679 LFCYFRRVTEKKPTGLVTFTRQ--SLEDFFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 736  
Db 195 IMHYFERFCSCVPIGIVSEFKKITAAPDAFWSKSDVSLCAFKVHSFGLIEDQPDNALEV 254  
QY 737 DFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNECLIITGTEQYSEYTGYA 796  
Db 255 DFANKYLGGSLSRGCVQEEIRFMINPELIAGMLFLPRMDONEAIEIVGAERFSCYTGYA 314  
QY 797 ETYRWSRSHEDGSDCERRCTEIVAIDAL-----HFRRYLDQFVPEKMRRELNKAYC 850  
Db 315 SSFRPAGEYIDKKAMPFKRRTRIVAIDALCTPKMRHFKDIC-----LLREINKALC 367  
QY 851 GFLR----- 854  
Db 368 GFLNCSKAWEHQNIFMDEGDNEIQLVRNGRDSGLLRTETTASHRTPLNDVEMNREKPA NN 427  
QY 855 -----PGVSSENL--SAVATGNWCGGAFGGDARLKALIQILAAAAERDVV-YFTFGD 904  
Db 428 LIRDFYVEGVNDHEDDGVATGNWCGGVFGGDPPELKATIQLWLAASQTRRRPFISYYTFG- 486  
QY 905 SELMRDIYSMHIFLTERKLTVDGVYKLLRY 935  
Db 487 VEALRNLDQVTKWILSHKWTVGDLWNMMLEY 517  
RESULT 12  
Q8VYA1  
ID Q8VYA1 PRELIMINARY; PRT; 522 AA.  
AC Q8VYA1;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)  
DE Putative poly (ADP-ribose) glycohydrolase.  
GN AT2G31870.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi s.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
RA Palm C.J., Bowser I., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AY072330; AAL61937.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;  
Query Match 10.4%; Score 539; DB 10; Length 522;  
Best Local Similarity 30.7%; Pred. No. 3.7e-30;  
Matches 133; Conservative 82; Mismatches 138; Indels 80; Gaps 11;  
QY 576 FWDKVLBEAQAHLYSILPDMVKIALCLPNICTQPIPLLKQKMNH----- 621  
Db 80 FFDKKISREESANFFGEVVPALCRLLQLPSMLEKH Y---QKADHVL DGVKSGRL L L GP 135  
QY 622 ----SITMSQEQLASLLANAFCTFPRNNAKMKSEYSSYPDINFNRLFE----GRSSRKP 673  
Db 136 QEAGIVLLSQELIAALLACSFCLFPFVDRSLK----NLQGINFSGLSFPYMRHCTKQE 191  
QY 674 EKLKTLFCYFRRVTEKKPTGLVTFTRQSL E-----DFPE---WERCEKPLTRLHVTY 722  
Db 192 NKIKCLIHYPGICRMWPTGFSFERKILPLEYHPHFVSPKADSWANSVTPLCSIEIHT 251

QY 723 EGTIEENGQGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNECLI 782  
Db 252 SGAIEDQPCEALEVDFADEYFGGLTLSYDTLQEEIRFVINPELIAGMIFLPRMDANEAI E 311  
QY 783 ITGTEQYSEYTGAEYRWSRSHEDGSDCERRCTEIVAIDALHFRRYLDQFVPEKMR 842  
Db 312 IVGVERFSGYTGYPSPFOYAGDYTDNKDLDFRRRKRTRVIAIDAMP-DPGMGQYKLDALI 370  
QY 843 RELNKAYCGFLR-----PGVSS-----ENLSAVA 866  
Db 371 REVNKAFSGYMHQCKYNIDVKHDPEASSSHVPLTSDSASQVIESSHRWCIDHEEKIGVA 430  
QY 867 TGNWCGGAFGGDARLKALIQILAAAAERDVV-YFTFGDSELMRDIYSMHIFLTERKLT V 925  
Db 431 TGNWCGGVFGGDPPELKIMQLWLAIQSQRPFMSYYTFG-LQALQNLNQVIEVMVALQEMTV 489  
RESULT 13  
Q867X0  
ID Q867X0 PRELIMINARY; PRT; 781 AA.  
AC Q867X0;  
DT 01-JUN-2003 (TremBLrel. 24, Created)  
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3 protein) (Corresponding sequence F20C5.1a).  
DE F20C5.1 OR PME-3.  
GN Caenorhabditis elegans.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode Caenorhabditis elegans.";  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";  
RT Science 282:2012-2018(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Matthews P.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY185493; AAO26316.1; -.  
DR EMBL; Z68161; CAD89735.1; -.  
DR WormPep; F20C5.1a; CE33775.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 781 AA; 89273 MW; 30447D06605CED0A CRC64;  
Query Match 9.3%; Score 480.5; DB 5; Length 781;  
Best Local Similarity 24.4%; Pred. No. 1.1e-25;  
Matches 215; Conservative 136; Mismatches 310; Indels 219; Gaps 39;  
QY 145 YLNQHQTAAMCKWQNEGKHTEQLLESEPTVTLVPEQFSNANIDRSPQNDHSDTDSEEN 204  
Db 27 FAHQVPTMKRRKLTGHGNTTESKED-----PEEPKSRDVFVSSQSSDESQEDSAEN 77  
QY 205 -----RDNQQFLT-TVKLANAKQ-----TTEDEHAREAKSHQKSKSCHPGEDCASCQ 252  
Db 78 PEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLD--NH----KSTEPMEE----- 125  
QY 253 DEIDVVPKSP LSDVGS EDVGTGSKNDKNKLI RQESCLGNSPPFEKESPESPMDVDNSKNS 312

Db 126 ---DVNKSNI-----DVAINSEDEDELVEEN-----NKE-----MRDGEQV 160

QY 313 QDSEADEE---TSPGFDEQ-----DGSSSQANKPSRFQARDADIEFRKRYSTKGGE 363

Db 161 QDQFADDDQELIEYFGIMKDTTQDITDSEVETAQKMEMIEETAD-----STFVGE 213

QY 364 VRLHFQFEGGESRTGMNDLNALPGNISLNEVECRNSQHGKDKSKITDHLMLPKAEDR 423

Db 214 DSKNQR-----QSGTTSDEVDADSQINLATKTVRTSSS-----SFLSTVSTCEAP 258

QY 424 RKEQWETGHQRTKIPKYPVPHLS--PDKKWLGTPIEMRRMPCRGIRLPLLRPSANHT 481

Db 259 AKGRARMYQKELEKHVIAFTGNTLQPD-----LNKVDPRNY- 297

QY 482 VTIRVDLLRAGEVPPKPPFTHYKDL-WDNKH---VKMPCSEONLYPVEDENGERTAGSRWE 537

Db 298 -----RYCTIPN-PPASQKLRDNRYGPKIVLP-----QRW- 328

QY 538 LIQTALLNKFTRPQNLKDALIKYN---VAYSKKWDFTA-----LIDFWD---KVLLEA 584

Db 329 -----REFDSRGRRRDSYFYFKRKLGDYKCYKTTGYFMFVGLLHNMWFEFDPITYKL 381

QY 585 EAQHLVQSILPDMV-----KIALCLPNICTQPIPLKQKMNHSITMSQEQIA 631

Db 382 PALEMYYKEMSELVGREVELEKFAVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCA 440

QY 632 SLLANAFCTFPRRNAMKSEYSSYPDINFRNLFEGRSRKPPEKLTFLCYFRRVTEKKP 691

Db 441 ALVARMFF-----ARPDSPFS-----FCRILSSDKSICVEKLFYFDKMSMDPP 487

QY 692 TGLVTF--TRQSLEDF-PEWERCCKPLTRLHVT--YEGTIEENGQGLQVDFANRFVGGG 746

Db 488 DGAVSFRLTQMDKDTFNEEWK--DKKLSLPEVEFFDEMLEDIALCTQVDFANEHLGGG 545

QY 747 VTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGEOYSEYTYGAETYRW----- 801

Db 546 VLNHGSVQEEIRFLMCPMMVMGMLLCEKMKQLEAISIVGAYVFPSSYTGHTLKWAELOP 605

QY 802 SRSHEDGSE-RDDCERRCTEIVAIDALHFR--YLDQFVPEKMRRELNKAYCGFLRP 855

Db 606 NHRQNTNEFRDRFGRRLRVETIAIDAILFKGSKLDCQTEQNLKANIIREMKKASIGFMSQ 665

QY 856 GVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDIYSMH 915

Db 666 GPKFTNI-PIVTGWGCGAFNGDKPLKFLQVIAAGVADRLPHFCSFGEPELAACKKII 724

QY 916 IFLTERKLTGVDVYKLLRLRYNEECRNCSTPGPDIKLYPF 955

Db 725 ERMKQKDVTLGMLFSMI-----NNTGLPHKHFEFYVF 756

RESULT 14

Q19637 PRELIMINARY; PRT; 764 AA.

AC Q19637;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose metabolism enzyme-3 short form).

GN F20C5.1 OR PME-3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Matthews P.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";

RL Science 282:2012-2018(1998).

RN [3]

RP SEQUENCE FROM N.A.

RA Gagnon S.N., Hardy I., Desnoyers S.;

RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode Caenorhabditis elegans.";

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z68161; CAA92299.2; --

DR EMBL; AY185494; AAC26317.1; --

DR PIR; T21138; T21138.

DR WormPep; F20C5.1b; CE32867.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

SQ SEQUENCE 764 AA; 87441 MW; B391703B1D0CB79A CRC64;

Query Match 9.1%; Score 473; DB 5; Length 764;

Best Local Similarity 23.6%; Pred. No. 3.5e-25;

Matches 211; Conservative 128; Mismatches 290; Indels 266; Gaps 38;

QY 145 YLNQHQTAAAMCKWQNEGKHTEQLLESEPQTVTLVPEQFSNANIDRSPQNDHSDTDSEEN 204

Db 27 FAHQVPTMKRRKLTEHGNTTESKED-----PEPKSRDVFVSSQSDSEQEDSAEN 77

QY 205 -----RDNQQLT-TVKLANAQ-----TTEDEHAREAKSHQCKSCHPCGEDCASCQQ 252

Db 78 PEIAKEVSENCENTETLKISNIESLDNVTERSEHTLD--NH-----KSTEPMEE----- 125

QY 253 DEIDVVPKSPLSVGSSEVDGTSKNDKNIQESCLGNSPPFEKESEPESPMDVDNSKNS 312

Db 126 ---DVNKSNI-----DVAINSEDEDELVEEN-----NKEMR-----DGEQVQQL 163

QY 313 QDSEADEE---TSPGFDEQ-----DGSSSQANKPSRFQARDADIEF-----RKR 356

Db 164 SQDLFADDDQELIEYFGIMKDTTQDITDSEVETAQKMEMIEETADSTFVGEDSKATKT 223

QY 357 YSTKG-----GEVRLH-----FQEGGESRTGMNDLNALPG----- 388

Db 224 VRTSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTG-NTLQPDINKVDPDRNRY 282

QY 389 ---NISLNEVECRNSQHGKDKSKITDHLMLPKAEDRRKEQWETKHORTE-----RK 438

Db 283 CTIPNFPASQKLRDNRYGPK-----IVLP-----QRWREFDSRGRRRDSYFYFKRK 330

QY 439 IPKVVPPLSPDKKWLGTPIEBMRM-PRCGIRLPLLRPSANHTVTIRVDLLRAGEVPKP 497

Db 331 LDGYLKCYKTTGYFMFVGLLHNMWFEFDPITYKLPALE----- 368

QY 498 FPTHYKDLWDNKHVKMPCSEONLYPVEDENGERTAGSRWELI-QTALLNKFTRPQNLKDA 556

Db 369 --MYKEM-----SELVGREVELEKFAVARIA-- 394

QY 557 ILKYNVAYSKKWDFTALIDFWDKVLLEEAQHLVQSILPDMVKIALCLPNICTQPIPLK 616

Db 395 -----KTAEDILPERIYR-LVGDV----- 412

QY 617 QKMNHSITMSQEQIASLIANAFCTFPRRNAMKSEYSSYPDINFRNLFEGRSRKPPEKL 676

Db 413 ----ESATLSHKQCAALVARMFF-----ARPDSPFS-----FCRILSSDKSICVEKL 455

QY 677 KTLFCYFRRVTEKKPTGLVTF--TRQSLEDF-PEWERCCKPLTRLHVT--YEGTIEENGQ 731

Db 456 KFLFTYFDKMSMDPPDGAVSFRLTQMDKDTFNEEWK--DKKLSLPEVEFFDEMLEDIAL 513

QY 732 GMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGEOYSE 791

Db 514 LCTQVDFANEHLGGVNLHGVSQEEIRFLMCPMMVMGMLLCEKMKQLEAISIVGAYVFS 573

QY 792 YTYGAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFR--YLDQFVPEK 840

Db 574 YTYGHTLKWAELOPNHSRQNTNEFRDRFGRRLRVETIAIDAILFKGSKLDCQTEQNLKAN 633

QY 841 MRRELKAYCYGLRPGVSSSENLSAVATGNWGCAGFGDARLKALIQILAAAAAERDVVYF 900  
Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGAFNGDKPLKFIQVIAAGVADRPPLHFC 692  
QY 901 TFGDSELMRDIYSMHIFLTERKLTGVDVYKLLRLRYNEECRNCSTPGPDIKLYPF 955  
Db 693 SFGEPELAACKKIIERMKQKDVTLGMLFSMI-----NNTGLPHKHFEFYVF 739

RESULT 15  
Q9N5L4 PRELIMINARY; PRT; 485 AA.  
AC Q9N5L4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme  
DE 4).  
GN H23L24.5 OR PME-4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Abu-Threideh J.; Lehnert L.;  
RT "The sequence of C. elegans cosmid H23L24.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Molecular characterization of poly ADP-ribose glycohydrolases in  
RT Caenorhabditis elegans.";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AC006662; AAF39896.2; -.  
DR EMBL; AF548468; AAN40699.1; -.  
DR WormPep; H23L24.5; CE32685.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;

Query Match 8.7%; Score 453; DB 5; Length 485;  
Best Local Similarity 30.7%; Pred. No. 5e-24;  
Matches 136; Conservative 74; Mismatches 157; Indels 76; Gaps 15;

QY 536 WELIQTALINKFTR-PQNLKDAILK--YNVAYS-----KKWDF TALIDF 576  
Db 46 FELLETGVSQQWRNCDQNLFEYLYKTKYKNGYSQFEDLLFKINGYSEKERFDLPALKSF 105  
QY 577 WDKVLE--EAAQHLVQSILPDMVKIALCLPNICTQPIPLLKQKMN-----HSITMSQ 627  
Db 106 YRKMSIEVGDE-----VLEKLARLVRIKTSAC---EVLPEKIYRLVGDIESATFSH 154  
QY 628 EQIASLIANAFFCTFPRNAKMKSEYSSYPDINFRNLFEGRSSRKPEKLTLCYFRRVT 687  
Db 155 IQCASLIAMWFFSDTFR-----LSFIILQKTTCAVEKLEKLFITYFDKMS 200

QY 688 EKKPTGLVTF-----RQSLDFPEWE-RCEKPLTRLHVTYEGTIEENGQGMQLQVDFAN 740  
Db 201 IDPPIGAVSFRKMRITHKQYLEN---WKLRETNLLPDVQVFDKMSIETAL-CTQIDFAN 256  
QY 741 RFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQYSEYTGZYAEYR 800  
Db 257 KRLGGVVKGGAVQEEIRFMCMPEMMVAILLNDVTQDLEAISIVGAYVFSSYTGYSNTLK 316  
QY 801 WSR-----SHEDGSRDDCERRCTEIVADALH-----FRYLDQFVPEKMRRELNKAY 849  
Db 317 WAKITPKHSAQNNSFRDQFGRLOQTETVAIDAVRNAGTPLECLLNQLTTEKLTREVRKAA 376  
QY 850 CGFLRPGVSSSENLSAVATGNWGCAGFGDARLKALIQILAAAAAERDVVYFTFGDSELMR 909  
Db 377 IGFLSAGDGFSKI-PVVGWGWGCGAFRGKPKLKFLIQVIACGISDRPLQFCTFGDTELA 435  
QY 910 DIYSMHIFLTERKLTGVDVYKLL 932  
Db 436 KCEEMMTLFRNNNVRTGQLFLII 458

Search completed: May 26, 2004, 18:45:59  
Job time : 81.8912 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 77.971 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-2  
Perfect score: 5184  
Sequence: 1 MSAGPGCEPCTKRPWDAAA.....YHAVESCTQTNPQRTGA 977

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB | ID     | Description         |
|------------|--------|-------------|--------|----|--------|---------------------|
| 1          | 5184   | 100.0       | 977    | 6  | 002776 | O02776 bos taurus   |
| 2          | 4694.5 | 90.6        | 976    | 4  | Q86W56 | Q86W56 homo sapien  |
| 3          | 4689.5 | 90.5        | 976    | 4  | Q7Z742 | Q7Z742 homo sapien  |
| 4          | 4635.5 | 89.4        | 976    | 4  | Q9Y4W7 | Q9Y4W7 homo sapien  |
| 5          | 4332.5 | 83.6        | 968    | 11 | O88622 | O88622 mus musculus |
| 6          | 4329   | 83.5        | 961    | 11 | Q80YQ6 | Q80YQ6 mus musculus |
| 7          | 4295.5 | 82.9        | 972    | 11 | Q9QYM2 | Q9qym2 rattus norv  |
| 8          | 4125   | 79.6        | 920    | 11 | Q8CB72 | Q8cb72 mus musculus |
| 9          | 1063.5 | 20.5        | 768    | 5  | O46043 | O46043 drosophila   |
| 10         | 1061.5 | 20.5        | 723    | 5  | Q960N8 | Q960n8 drosophila   |
| 11         | 585    | 11.3        | 548    | 10 | Q9SKB3 | Q9skb3 arabidopsis  |
| 12         | 533    | 10.3        | 522    | 10 | Q8VYA1 | Q8vya1 arabidopsis  |
| 13         | 481.5  | 9.3         | 781    | 5  | Q867X0 | Q867x0 caenorhabdi  |
| 14         | 475    | 9.2         | 764    | 5  | Q19637 | Q19637 caenorhabdi  |
| 15         | 472    | 9.1         | 485    | 5  | Q9N5L4 | Q9n5l4 caenorhabdi  |
| 16         | 335.5  | 6.5         | 368    | 5  | Q86GI4 | Q86gi4 toxoplasma   |

|    |       |     |      |    |        |                    |
|----|-------|-----|------|----|--------|--------------------|
| 17 | 331   | 6.4 | 364  | 10 | Q9SKB4 | Q9skb4 arabidopsis |
| 18 | 185   | 3.6 | 508  | 12 | Q9E234 | Q9e234 helicoverpa |
| 19 | 181   | 3.5 | 571  | 5  | Q8MTN7 | Q8mtn7 trichinella |
| 20 | 180   | 3.5 | 510  | 12 | Q99GU9 | Q99gu9 helicoverpa |
| 21 | 180   | 3.5 | 2081 | 10 | Q9LH98 | Q9lh98 arabidopsis |
| 22 | 179.5 | 3.5 | 1078 | 5  | Q963T1 | Q963t1 plasmodium  |
| 23 | 171.5 | 3.3 | 1444 | 5  | Q9VTN2 | Q9vtn2 drosophila  |
| 24 | 171.5 | 3.3 | 1514 | 5  | Q8SY55 | Q8sy55 drosophila  |
| 25 | 171   | 3.3 | 791  | 13 | Q9DGL1 | Q9dgl1 fugu rubrip |
| 26 | 169.5 | 3.3 | 678  | 13 | Q803F8 | Q803f8 brachydanio |
| 27 | 167   | 3.2 | 455  | 5  | Q86A18 | Q86a18 dictyosteli |
| 28 | 165   | 3.2 | 968  | 5  | Q8I5W9 | Q8i5w9 plasmodium  |
| 29 | 164.5 | 3.2 | 1012 | 5  | Q8SXP2 | Q8sxp2 drosophila  |
| 30 | 164.5 | 3.2 | 1330 | 5  | Q9VJ87 | Q9vj87 drosophila  |
| 31 | 164   | 3.2 | 1115 | 5  | Q17863 | Q17863 caenorhabdi |
| 32 | 164   | 3.2 | 1618 | 16 | Q7UYM1 | Q7uym1 rhodopirell |
| 33 | 163.5 | 3.2 | 498  | 5  | Q8MTN8 | Q8mtn8 trichinella |
| 34 | 162.5 | 3.1 | 1071 | 10 | Q9M0D9 | Q9m0d9 arabidopsis |
| 35 | 162.5 | 3.1 | 7210 | 5  | Q9V7G8 | Q9v7g8 drosophila  |
| 36 | 162.5 | 3.1 | 9270 | 5  | Q8MLD9 | Q8mld9 drosophila  |
| 37 | 161.5 | 3.1 | 3398 | 13 | Q8QFV6 | Q8qfv6 gallus gall |
| 38 | 161   | 3.1 | 691  | 4  | Q96QF7 | Q96qf7 homo sapien |
| 39 | 160.5 | 3.1 | 1152 | 4  | Q92603 | Q92603 homo sapien |
| 40 | 160.5 | 3.1 | 1204 | 4  | O00211 | O00211 homo sapien |
| 41 | 159.5 | 3.1 | 1156 | 5  | Q8IM05 | Q8im05 plasmodium  |
| 42 | 159.5 | 3.1 | 3111 | 5  | Q9VH10 | Q9vhl0 drosophila  |
| 43 | 159   | 3.1 | 1535 | 10 | Q9LXZ9 | Q9lxx9 arabidopsis |
| 44 | 159   | 3.1 | 1817 | 4  | O43166 | O43l66 homo sapien |
| 45 | 159   | 3.1 | 8081 | 5  | Q7Z120 | Q7z120 caenorhabdi |

ALIGNMENTS

RESULT 1

O02776 PRELIMINARY; PRT; 977 AA.  
AC O02776;  
DT 01-JUL-1997 (Tremblrel. 04, Created)  
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN BPARG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9277328; PubMed=9115250;  
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase."  
RL J. Biol. Chem. 272:11895-11901(1997).  
DR EMBL; U78975; AAB53370.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
DR KW Hydrolase.  
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;

Query Match 100.0%; Score 5184; DB 6; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |    |   |
|----|----|---|
| QY | 1  | MSAGPGCEPCTKRPWDAAATSPPAASDARSFPGRRVLDSDAPVQFRVPPSSSGCAL 60     |
| Db | 1  | MSAGPGCEPCTKRPWDAAATSPPAASDARSFPGRRVLDSDAPVQFRVPPSSSGCAL 60     |
| QY | 61 | GRAGQHRGATSIVFKQKTTITWMDTKIKTVESESLHSESKENNTREESMMSSVQKDNFY 120 |
| Db | 61 | GRAGQHRGATSIVFKQKTTITWMDTKIKTVESESLHSESKENNTREESMMSSVQKDNFY 120 |

|    |   |  |
|----|---|--|
| OX | NCBI_TaxID=9606;  |  |
| RN | [1]   |  |
| RP | SEQUENCE FROM N.A.  |  |
| RC | TISSUE=Skin;  |  |
| RA | Strausberg R.;  |  |
| RL | Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.               |  |
| DR | EMBL; BC050560; AAH50560.1; -   |  |
| DR | GO; GO:0016787; F:hydrolase activity; IEA.                            |  |
| DR | InterPro; IPR007724; PARG.  |  |
| DR | Pfam; PF05028; PARG; 1.   |  |
| KW | Hydrolase.  |  |
| SQ | SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;                   |  |
|    | Query Match 90.6%; Score 4694.5; DB 4; Length 976;                    |  |
|    | Best Local Similarity 90.5%; Pred. No. 0;                             |  |
|    | Matches 883; Conservative 29; Mismatches 63; Indels 1; Gaps 1;        |  |
| QY | 1 MSAGPGCEPCTKRPRWDAATSPPAASDARSFPGRRVLDSDKAPVQFRVPPSSSGCAL 60        |  |
| Db | 1 MNAGPGCEPCTKRPRWGAATTS -PAASDARSFPGRRVLDPKDAHVQFRVPPSSPACVP 59      |  |
| QY | 61 GRAGQHRGATSIVFKQKTTITSWMDTKGKTVSESLHSENNNTREESNMSSVQKDNFY 120      |  |
| Db | 60 GRAGQHRGATSIVFKQKTTITSWMDTKGKTAESLSDSKENNNTRIESNMSSVQKDNFY 119     |  |
| QY | 121 QHNMEKLENVSQLGPDKSPVEKGTQYLKQHOTAAMCKWQNEGPHSERLLESEPPAVTLVP 180  |  |
| Db | 120 QHNVEKLENVSQLSLDKSPTEKSTQYLNQHOTAAMCKWQNEGKHTQLLESEPTVTLVP 179    |  |
| QY | 181 EQFSNANVDQSSPKDDHSDTNSSESRDQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240    |  |
| Db | 180 EQFSNANIDRSPQNDHSDTDSSENRDQFLTHVKLANAKQTTEDEQAREAKSHQKCSK 239     |  |
| QY | 241 ACHPAEACAGCQCEETDVVSESPSLDTGSEVDVGTGLKNANRLNRQESSLGNPPFEKSE 300   |  |
| Db | 240 SCDPGEDECASCQDEIDVVPESPLSDVGSVDVGTGPKNDNKLTRQESCLGNPPFEKSE 299    |  |
| QY | 301 PESPMVDVNSKNSCQDSEADETSPGFDEQEDSSSAQTANKPSRFQPREADTELKRSSA 360    |  |
| Db | 300 PESPMVDVNSKNSCQDSEADETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKYST 359    |  |
| QY | 361 KGGEIRLHGFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGKDKSKITDHFMRVPK 420 |  |
| Db | 360 KGGEVRLHGFQFEGGESRTGMDLNAKLPGNISLNVECRNSKQHGKDKSKITDHFMRLPK 419   |  |
| QY | 421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPCGIRLPPLRPSAN 480   |  |
| Db | 420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPPDKKWLGTPIEEMRRMPCGIRLPPLRPSAN 479  |  |
| QY | 481 HTVTIRVDLLRIGEVKPPFTHFKDLWDNKHVKMPCSEONLYPVEDENGERAAGSRWELI 540   |  |
| Db | 480 HTVTIRVDLLRAGEVPKPFTHYKDLWDNKHVKMPCSEONLYPVEDENGERTAGSRWELI 539   |  |
| QY | 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFLTALIDFWDKVLSEAEAAHLYQSILPDMVK 600 |  |
| Db | 540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFLTALIDFWDKVLSEAEAAHLYQSILPDMVK 599 |  |
| QY | 601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660    |  |
| Db | 600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659    |  |
| QY | 661 NFNRLFEGRSRRKPEKLTLCFVFRVTEKKPTGLVTFTTQSLDFFPEWERCEKLLTRLH 720    |  |
| Db | 660 NFNRLFEGRSRRKPEKLTLCFVFRVTEKKPTGLVTFTTQSLDFFPEWERCEKPLTRLH 719    |  |
| QY | 721 VTYEGTIRGNGQGMLQVDVFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTFVLVDHNE 780 |  |
| Db | 720 VTYEGTIRGNGQGMLQVDVFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTFVLVDHNE 779 |  |
| QY | 781 CLIIITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRTEIIVADALHFRYLDQFVPE 840   |  |
| Db | 780 CLIIITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRCTEIVADALHFRYLDQFVPE 839   |  |

|    |  |  |
|----|--|--|
| QY | 121 QHNMEKLENVSQLGPDKSPVEKGTQYLKQHOTAAMCKWQNEGPHSERLLESEPPAVTLVP 180   |  |
| Db | 121 QHNMEKLENVSQLGPDKSPVEKGTQYLKQHOTAAMCKWQNEGPHSERLLESEPPAVTLVP 180   |  |
| QY | 181 EQFSNANVDQSSPKDDHSDTNSSESRDQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240     |  |
| Db | 181 EQFSNANVDQSSPKDDHSDTNSSESRDQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240     |  |
| QY | 241 ACHPAEACAGCQCEETDVVSESPSLDTGSEVDVGTGLKNANRLNRQESSLGNPPFEKSE 300    |  |
| Db | 241 ACHPAEACAGCQCEETDVVSESPSLDTGSEVDVGTGLKNANRLNRQESSLGNPPFEKSE 300    |  |
| QY | 301 PESPMVDVNSKNSCQDSEADETSPGFDEQEDSSSAQTANKPSRFQPREADTELKRSSA 360     |  |
| Db | 301 PESPMVDVNSKNSCQDSEADETSPGFDEQEDSSSAQTANKPSRFQPREADTELKRSSA 360     |  |
| QY | 361 KGGEIRLHGFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGKDKSKITDHFMRVPK 420  |  |
| Db | 361 KGGEIRLHGFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGKDKSKITDHFMRVPK 420  |  |
| QY | 421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPCGIRLPPLRPSAN 480    |  |
| Db | 421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPCGIRLPPLRPSAN 480    |  |
| QY | 481 HTVTIRVDLLRIGEVKPPFTHFKDLWDNKHVKMPCSEONLYPVEDENGERAAGSRWELI 540    |  |
| Db | 481 HTVTIRVDLLRIGEVKPPFTHFKDLWDNKHVKMPCSEONLYPVEDENGERAAGSRWELI 540    |  |
| QY | 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFLTALIDFWDKVLSEAEAAHLYQSILPDMVK 600  |  |
| Db | 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFLTALIDFWDKVLSEAEAAHLYQSILPDMVK 600  |  |
| QY | 601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660     |  |
| Db | 601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660     |  |
| QY | 661 NFNRLFEGRSRRKPEKLTLCFVFRVTEKKPTGLVTFTTQSLDFFPEWERCEKLLTRLH 720     |  |
| Db | 661 NFNRLFEGRSRRKPEKLTLCFVFRVTEKKPTGLVTFTTQSLDFFPEWERCEKLLTRLH 720     |  |
| QY | 721 VTYEGTIRGNGQGMLQVDVFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTFVLVDHNE 780  |  |
| Db | 721 VTYEGTIRGNGQGMLQVDVFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTFVLVDHNE 780  |  |
| QY | 781 CLIIITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRTEIIVADALHFRYLDQFVPE 840    |  |
| Db | 781 CLIIITGTEQYSEYTGAEITYRWARSHEDRSERDDWQRRTEIIVADALHFRYLDQFVPE 840    |  |
| QY | 841 KIRRELNKAYCGFLRPGVSSSENLSAVATGNWCGAGFGGDARLKALIQILAAVAERDVVY 900   |  |
| Db | 841 KIRRELNKAYCGFLRPGVSSSENLSAVATGNWCGAGFGGDARLKALIQILAAVAERDVVY 900   |  |
| QY | 901 FTFGDSSELMRDIYSMHTFLTERKLTGVEVYKLLRLRYNNEECRNCSTPGPDIKLYPFIYHA 960 |  |
| Db | 901 FTFGDSSELMRDIYSMHTFLTERKLTGVEVYKLLRLRYNNEECRNCSTPGPDIKLYPFIYHA 960 |  |
| QY | 961 VESCTQTTNQPGQRTGA 977  |  |
| Db | 961 VESCTQTTNQPGQRTGA 977  |  |

RESULT 2

Q86W56

ID Q86W56 PRELIMINARY; PRT; 976 AA.

AC Q86W56;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Similar to poly (ADP-ribose) glycohydrolase.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

|    |     |   |     |
|----|-----|---|-----|
| QY | 841 | KIRRELKAYCGFLRPGVSSSENLSAVATGNWGCAGFGDARLKALIQILAAVAERDVVY  | 900 |
| Db |     | : |     |

|          |   |   |              |
|----------|---|---|--------------|
| QY       | 181   | EQFSNANVDQSSPKDDHSDTNSERSDNQOFLTHVKLANAKOTMEDEQGREARSHQCKG    | 240          |
| Db       | 180   | EQFSNANIDRSPQNDHSDTSEENRDNQOFLTTVKLANAKOTTEDEQAREAKSHQCKSK    | 239          |
| QY       | 241   | ACHPAEACAGCQEEETDVVSESPISDGTGSEDEVGTGLKNANRLNRQESSLGNSPPFKEKE | 300          |
| Db       | 240   | SCDPGEDCASCQDEIDVVPESPISDVGSEDEVGTGPKNDKNLTKRQESCLGNSPPFKEKE  | 299          |
| QY       | 301   | PESPMVDNSKNQSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELKRRSSA  | 360          |
| Db       | 300   | PESPMVDNSKNQSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST  | 359          |
| QY       | 361   | KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK   | 420          |
| Db       | 360   | KGGEVRLHFQFEGGESRTGMNDLNKLPGNISLNVECRNSKHGKDSKITDHFMRLPK      | 419          |
| QY       | 421   | AEDKRKEQCEMKHQRTERKIPKYIPHLSPDKKWLGTPIEMRRMRPRCGIRLPPLRPSAN   | 480          |
| Db       | 420   | AEDRRKEQWETHQRTERKIPKYVPPHLSPDKKWLGTPIEMRRMRPRCGIRLPPLRPSAN   | 479          |
| QY       | 481   | HTVTIRVDLLRIGEVKPPPTFKDLNDKNHVKMPCSEONLYPVEDENGERAAGSRWELI    | 540          |
| Db       | 480   | HTVTIRVDLLRAGEVPKPPPTHYKDLNDKNHVKMPCSEONLYPVEDENGERTAGSRWELI  | 539          |
| QY       | 541   | QTALLNRLTRPQNLKDAILKYNVAYSKKWDFALIDFWDKVLSEAAQHLKQSIPLDMVK    | 600          |
| Db       | 540   | QTALLNKFTRPQNLKDAILKYNVAYSKKWDFALIDFWDKVLSEAAQHLKQSIPLDMVK    | 599          |
| QY       | 601   | IALCLPNICTQPIPLLKQKNHSITMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDI    | 660          |
| Db       | 600   | IALCLPNICTQPIPLLKQKNHSITMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDI    | 659          |
| QY       | 661   | NFNRLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTQSLDFPEWERCEKLLTRLH      | 720          |
| Db       | 660   | NFNRLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTQSLDFPEWERCEKPLTRLH      | 719          |
| QY       | 721   | VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEHIFRLINPELIVSRLEFTEVLHDNE  | 780          |
| Db       | 720   | VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQEEHIFRLINPELIIISRLFTEVLHDNE  | 779          |
| QY       | 781   | CLIIITGEQYSEYTGABTYRWARSHEDRSERDDWQRRTEIIVADALHFRRYLDQFVPE    | 840          |
| Db       | 780   | CLIIITGEQYSEYTGABTYRWSRSHEDGSRDDWQRRHCTEIVADALHFRRYLDQFVPE    | 839          |
| QY       | 841   | KIRRELKAYCGFLRPGVSSSENLSAVATGNWGCAGFGDARLKALIQILAAVAERDVVY    | 900          |
| Db       | 840   | KMRRELKAYCGFLRPGVSSSENLSAVATGNWGCAGFGDARLKALIQILAAAAERDVVY    | 899          |
| QY       | 901   | FTFGDSELMRDIYSMHTFLTERKLTGVEVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHA  | 960          |
| Db       | 900   | FTFGDSELMRDIYSMHIFLTERKLTGVDVYKLLRLRYNEECRNCSTPGPDIKLYPFIYHA  | 959          |
| QY       | 961   | VESCTQTNNQPGQRTG  | 976          |
| Db       | 960   | VESCAETADHSGQRTG  | 975          |
| RESULT 4 |   |   |              |
| ID       | Q9Y4W7  | PRELIMINARY;  | PRT; 976 AA. |
| AC       | Q9Y4W7;   |   |              |
| DT       | 01-NOV-1999   | (TREMBLrel. 12, Created)                                      |              |
| DT       | 01-NOV-1999   | (TREMBLrel. 12, Last sequence update)                         |              |
| DT       | 01-JUN-2003   | (TREMBLrel. 24, Last annotation update)                       |              |
| DE       | Poly(ADP-ribose) glycohydrolase.                                  |   |              |
| GN       | HPARG.  |   |              |
| OS       | Homo sapiens (Human).   |   |              |
| OC       | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |   |              |
| OC       | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.        |   |              |
| OX       | NCBI_TaxID=9606;  |   |              |
| RN       | [1]   |   |              |
| RP       | SEQUENCE FROM N.A.  |   |              |







Db 353 KGSEVRLHFQFE - GENNAGTSDLNAKPSGNSSSLNVECRSSKHQKGRDSKITDHFMRISK 411  
QY 421 AEDKRKEQCEMKHQTERKIPKIYIPPHLSPOKWLGTPIEBMRMRPCGIRLPLRPSAN 480  
Db 412 SEDRRKEQCEVHRQTERKIPKIYIPPNLPPEKKWLGTPIEBMRMRPCGIRLPLRPSAS 471  
QY 481 HTVTIRVDLLRIGEVKPPFTHFQKLDNDKHNKVMPCSEQNLYPVEDENGERRAAGSRWELI 540  
Db 472 HTVTIRVDLLRAGEVPPFTTHYKLDNDKHNKVMPCSEQNLYPVEDENGERTAGSRWELI 531  
QY 541 QTALLNRLTRPQNLKDALILKYNVAYSKKWDFTALIDFWDKVLAEAAQHLYSILPDMVK 600  
Db 532 QTALLNKFTRPQNLKDALILKYNVAYSKKWDFTALVDFWDKVLAEAAQHLYSILPDMVK 591  
QY 601 IALCLPNICTOPIPLLKQKMHSTMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYPDI 660  
Db 592 IALCLPNICTOPIPLLKQKMHSTMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYPDI 651  
QY 661 NFNLFEGRSSRKPEKLTILFCYFRRVTEKKPTGLVTFTRQSLEDFFPEWERCEKLLTRLH 720  
Db 652 NFNLFEGRSSRKPEKLTILFCYFRRVTEKKPTGLVTFTRQSLEDFFPEWERCEKPLTRLH 711  
QY 721 VTYEGTIEGNQGMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780  
Db 712 VTYEGTIEGNRGRMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771  
QY 781 CLIIITGEQYSEYTGAEYRWARSHEDRSERDDWQRRTEIIVAIIDALHFRYLDQFVPE 840  
Db 772 CLIIITGEQYSEYTGAEYRWARSHEDGSEKDDWQRRCTEIVAIIDALHFRYLDQFVPE 831  
QY 841 KIRRELNKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
Db 832 KVRRELNKAYCGFLRPGVPSNLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 891  
QY 901 FTFGDSELMDIYSMHTFLTERKLTVCGEVYKLLRYYNEECRCNSTPGPDIKLYPFIYHA 960  
Db 892 FTFGDSELMDIYSMHTFLTERKLDVGKVKLLRYYNEECRCNSTPGPDIKLYPFIYHA 951  
QY 961 VESCTQTTNPGQRTG 976  
Db 952 VESSAETDMPGQKAG 967

RESULT 6  
Q80YQ6 PRELIMINARY; PRT; 961 AA.  
AC Q80YQ6;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050892; AAH50892.1; -  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;  
Query Match 83.5%; Score 4329; DB 11; Length 961;  
Best Local Similarity 84.4%; Pred. No. 1.8e-303;  
Matches 818; Conservative 55; Mismatches 88; Indels 8; Gaps 4;  
QY 1 MSAGPGCEPCTKRPWDAAATSPPAASDARSFPGQRRLVDSKDAPVQVRPVPSSGGAL 60

Db 1 MSAGPGWEPCTKRPRWGAAGTSAPTASDSRSFPGQRRLVDPKAPVQVRPVPSSPACVS 60  
QY 61 GRAGQHRGSATSLVFKQKTITSMWDTKGIKTVESESLHSKENNTREESMMSSVQKDNFY 120  
Db 61 GRAGPHRGNATSFVFKQKTITSMWDTKGPKEASE - - - SKENNTRIDSMSSVQKDNFY 117  
QY 121 QHNMEKLENVSQJGFDKSPVEKGTQYLKQHQTAAMCKQWNEGPHSERLLESEPPAVTLVP 180  
Db 118 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 177  
QY 181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQLTHVVKLANAKQIMEDEQGREAPSHQKCGK 240  
Db 178 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQLTPIKLANTKPTVGDGQ - - - ARSNCKCSG 234  
QY 241 ACHPAEACAGCQOEETDVVSESPSLSDTGSSEVGTGLKNANRLNRQESSLGNSPPEKESE 300  
Db 235 SRQSVKDCGTGQQEEDVLPESPLSDVGAEDIGTGPKNDNKLITGQESSLGDSPPEKESE 294  
QY 301 PESPMVDNKNKSCQDSEADEETSPGFEDEQEDSSSAQTANKPSRFQPREADTELKRKSSA 360  
Db 295 PESPMVDNKNKSCQDSEADEETSPVFEDEQDRSS - QTANKLSSCQAREADGDLRKRYLT 353  
QY 361 KGGEIRLHFQFEGGESRAGMNDVNNAKPGSTSSLNEVCNRSKQHGKDKSKITDHFMRVVK 420  
Db 354 KGSEVRLHFQFE - GENNVGTSDLNAKPSGNSSSLNVECRSSKHQKGRDSKITDHFMRISK 412  
QY 421 AEDKRKEQCEMKHQTERKIPKIYIPPHLSPOKWLGTPIEBMRMRPCGIRLPLRPSAN 480  
Db 413 SEDRRKEQCEVHRQTERKIPKIYIPPNLPPEKKWLGTPIEBMRMRPCGIRLPLRPSAS 472  
QY 481 HTVTIRVDLLRIGEVKPPFTHFQKLDNDKHNKVMPCSEQNLYPVEDENGERRAAGSRWELI 540  
Db 473 HTVTIRVDLLRAGEVPPFTTHYKLDNDKHNKVMPCSEQNLYPVEDENGERTAGSRWELI 532  
QY 541 QTALLNRLTRPQNLKDALILKYNVAYSKKWDFTALIDFWDKVLAEAAQHLYSILPDMVK 600  
Db 533 QTALLNKFTRPQNLKDALILKYNVAYSKKWDFTALVDFWDKVLAEAAQHLYSILPDMVK 592  
QY 601 IALCLPNICTOPIPLLKQKMHSTMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYPDI 660  
Db 593 IALCLPNICTOPIPLLKQKMHSTMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYPDI 652  
QY 661 NFNLFEGRSSRKPEKLTILFCYFRRVTEKKPTGLVTFTRQSLEDFFPEWERCEKLLTRLH 720  
Db 653 NFNLFEGRSSRKPEKLTILFCYFRRVTEKKPTGLVTFTRQSLEDFFPEWERCEKPLTRLH 712  
QY 721 VTYEGTIEGNQGMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780  
Db 713 VTYEGTIEGNRGRMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 772  
QY 781 CLIIITGEQYSEYTGAEYRWARSHEDRSERDDWQRRTEIIVAIIDALHFRYLDQFVPE 840  
Db 773 CLIIITGEQYSEYTGAEYRWARSHEDGSEKDDWQRRCTEIVAIIDALHFRYLDQFVPE 832  
QY 841 KIRRELNKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900  
Db 833 KVRRELNKAYCGFLRPGVPSNLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 892  
QY 901 FTFGDSELMDIYSMHTFLTERKLTVCGEVYKLLRYYNEECRCNSTPGPDIKLYPFIYHA 960  
Db 893 FTFGDSELMDIYSMHTFLTERKLDVGKVKLLRYYNEECRCNSTPGPDIKLYPFIYHA 952  
QY 961 VESCTQTTN 969  
Db 953 VESSAETTD 961  
RESULT 7  
Q9QYM2 PRELIMINARY; PRT; 972 AA.  
AC Q9QYM2;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)

01-MAY-2000 (Tremblrel. 13, Last sequence update)  
01-JUN-2003 (Tremblrel. 24, Last annotation update)  
Poly(ADP-ribose) glycohydrolase.  
PARG.  
Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BUF; TISSUE=Colon;  
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,  
RA Aoki Y., Nakagawa H., Sugimura T.;  
RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase  
RT (Parg).";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB019366; BAA87901.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;  
Query Match 82.9%; Score 4295.5; DB 11; Length 972;  
Best Local Similarity 82.6%; Pred. No. 4.9e-301;  
Matches 808; Conservative 60; Mismatches 101; Indels 9; Gaps 4;  
QY 1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRRVLDSDAPVQFRVPPSSSGCAL 60  
Db 1 MSAGPGCEPCTKRPRWGAAAGTSAPTASDSRSFPGRRVLDSDAPVQFRVPPSSACVS 60  
QY 61 GRAGQHRGATSLSVFKQKTTTSMWDTKGITVSESLHSENNNTREESMMSSVQKDNFY 120  
Db 61 GRAGPHRGVTSFVFKQKPTTWMDTKGPKTAESE---SKENNTRTDPMSSVQKDNFY 117  
QY 121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHOTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
Db 118 PHKVEKLGNVLPQLNDKSPTEKSTPYLNQQQTAGVCKWHSAGERAELSAEPPSAVTQAP 177  
QY 181 EQFSNANVDQSSPKDDHSDTNSSESRDNOQFLTHVKLANAKQTMEDQGREARSHQKCGK 240  
Db 178 KQLSNANIDQSPPTDGHSDTDHEEDRDNOQFLTPVKLANAKQTVGDGQ---ARSNCKGSA 234  
QY 241 ACHPAEACAGCQCEETDVVSESPLSDTGSSEDVGTGLKNANLNQESSLGNSSPPFEKSE 300  
Db 235 SCQCGQDCAGCQCEADVPIESPPLSDVGAEDIGTGSNDNKLTKQESGLGDSPPFEKSE 294  
QY 301 PESPMVDVNSKNSQDSEADETSPGFEQEDSS--SAQTANKPSRFQPREADTELKRS 358  
Db 295 PESPMVDVNSKTSQDSEADEEASPVFDEQDDQDRSSQTANKLSRQAREVDGLRKRY 354  
QY 359 SAKGGEIRLHFOFEGGESRAGMNDVNAKRPSTSSLNVECRNSKQGRKDSKITDHFMRV 418  
Db 355 LTKGSEIRLHFOFEGG-SNAGTSDLNAPSGNSSSSLNVDGRSSKQHGKRDKITDHFVRI 413  
QY 419 PKAEDKRKEQCEMKHQRTERKIPKVIPLHSPDKKWLGTPIEEMRMPCGIRLPLRPS 478  
Db 414 PKSEDKRKEQCEVHQRAERKIPKYVPPNLPDPKKWLGTPIEEMRMPCGIRLPLRPS 473  
QY 479 ANHTVTIRVDLLRIGEPKPPPTHFKDLWDNKHVKMPCSEQLNYPVEDENGERAAGSRWE 538  
Db 474 ASHTVTIRVDLLRAGEVPKPPPTHYKDLWDNKHVKMPCSEQLNYPVEDENGERTAGSRWE 533  
QY 539 LIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFDALDQVLDKVLAEAAQHLQSIPLPDM 598  
Db 534 LIQTALLNKTTRPQNLKDAILKYNVAYSKKWDFDALDQVLDKVLAEAAQHLQSIPLPDM 593  
QY 599 VKIALCLPNICTQPIPLLKQNMHSITMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYP 658  
Db 594 VKIALCLPNICTQPIPLLKQNMHSITMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYP 653  
QY 659 DINFNRLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDFFPEWERCEKLTTR 718

654 DINFNRLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDFFPEWERCEKLTTR 713  
QY 719 LHVTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQEHIRFLINPELIVSRLFTVLDH 778  
Db 714 LHVTYEGTIEGNGRGLQVDFANFRVGGVTSAGLVQEHIRFLINPELIVSRLFTVLDH 773  
QY 779 NECLITGTEQYSEYTGATYRWARSHEDRSERDDWQRRTEIIVADALHERRYLDQFV 838  
Db 774 NECLITGTEQYSEYTGATYRWARSHEDGSEKDDWQRCCTEIVADALHERRYLDQFV 833  
QY 839 PEKIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAVAERDV 898  
Db 834 PEKIRRELKAYCGFLRPGVPPNLSAVATGNWCGAFGGDARLKALIQILAAVAERDV 893  
QY 899 VYFTFGDSELMRDIYSMHTFLTERKLTVEVYKLLRLRYNEECNCSTPGPDIKLYPFIY 958  
Db 894 VYFTFGDSELMRDIYSMHTFLTERKLVNKGKVRLLRLRYREECRDCSSPGPDIKLYPFIY 953  
QY 959 HAVESCTQTTNQPGQRTG 976  
Db 954 HAAESSAETSDDPGQRTG 971  
RESULT 8  
Q8CB72 PRELIMINARY; PRT; 920 AA.  
ID Q8CB72  
AC Q8CB72  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Poly.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;  
Query Match 79.6%; Score 4125; DB 11; Length 920;  
Best Local Similarity 84.3%; Pred. No. 9.1e-289;  
Matches 782; Conservative 52; Mismatches 86; Indels 8; Gaps 4;  
QY 1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRRVLDSDAPVQFRVPPSSSGCAL 60  
Db 1 MSAGPGWEPTKRPRWGAAAGTSAPTASDSRSFPGRRVLDSDAPVQFRVPPSSPACVS 60  
QY 61 GRAGQHRGATSLSVFKQKTTTSMWDTKGITVSESLHSENNNTREESMMSSVQKDNFY 120  
Db 61 GRAGPHRGVTSFVFKQKPTTWMDTKGPKTAESE---SKENNTRTDPMSSVQKDNFY 117  
QY 121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHOTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
Db 118 PHKVEKLGNVLPQLNDKSPTEKSSQYLNQQQTASVCKWQNEGKHAELASEPPAGTPLP 177  
QY 181 EQFSNANVDQSSPKDDHSDTNSSESRDNOQFLTHVKLANAKQTMEDQGREARSHQKCGK 240  
Db 178 KQLSNANIGQSPHTDDHSDTDHEEDRDNOQFLTPVKLANAKQTVGDGQ---ARSNCKGSA 234  
QY 241 ACHPAEACAGCQCEETDVVSESPLSDTGSSEDVGTGLKNANLNQESSLGNSSPPFEKSE 300

Db 235 SRQSVKDTGCGQEEVDVLPESPLSDVGAEDIGTGPKNDKNTLGTQESSLGDSPPFKESE 294

QY 301 PESPMVDNSKNSQDSEADETSPGFDEQEDSSAQTANKPSRFQPREADTELKRSSA 360

Db 295 PESPMVDNSKNSQDSEADETSPVFEQDDRSS-QTANKLSSCQAREADGLRKRYLT 353

QY 361 KGGEIRLHFQEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGKDSKITDHFVRPK 420

Db 354 KGSEVRLHFQFE-GENNAGTSDLNAPKPSGSSSLNVECRSSKQHGKDSKITDHFMRISK 412

QY 421 AEDKRKEQCEMKHQRTKIPKYIPPHLSPDKKWLGTPIEEMRRMPCGIRLPLRPSAN 480

Db 413 SEDRRKEQCEVRHQRTKIPKYIPPNLPPEKKWLGTPIEEMRMKPCGIRLPLRPSAS 472

QY 481 HTVTIRVDLLRIGEVPKPPPTHFKDLWDNKHVKMPCSEONLYPVEDENGERAAGSRWELI 540

Db 473 HTVTIRVDLLRAGEVPKPPPTHYKDLWDNKHVKMPCSEONLYPVEDENGERTAGSRWELI 532

QY 541 QTALLNRLTRPQNLKDAILLKYNVAYSKKWDTALIDFWDKVLSEAEQAHLQSIPLDMVK 600

Db 533 QTALLNKFTRPQNLKDAILLKYNVAYSKKWDTALVDVFDKVLSEAEQAHLQSIPLDMVK 592

QY 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYPDI 660

Db 593 IALCLPNICTQPIPLLKQKNHSITMSQEQIASLLANAFCTFPPRRNAKMKSEYSSYPDI 652

QY 661 NFNRLFEGRSSRKPEKLTLCFYFRRVTEKPTGLVTFTRQSLDFPEWERCEKLLTRLH 720

Db 653 NFNRLFEGRSSRKPEKLTLCFYFRRVTEKPTGLVTFTRQSLDFPEWERCEKPLTRLH 712

QY 721 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELIVSRFLTEVLDHNE 780

Db 713 VTYEGTIEGNGRGLQVDFANRFVGGVTSAGLVQBEIRFLINPELIVSRFLTEVLDHNE 772

QY 781 CLIIITGEQYSEYTGVAETRWARSHEDRSERDDWQRTTEIVAIDALHFRRYLDQFVPE 840

Db 773 CLIIITGEQYSEYTGVAETRWARSHEDGSEKDDWQRTTEIVAIDALHFRRYLDQFVPE 832

QY 841 KIRRELNKAYCGFLRPGVSSSENLSAVATGNWCGAFGQDARLKALIQILAAVAERDVVY 900

Db 833 KVRRELNKAYCGFLRPGVPSSENLSAVATGNWCGAFGQDVRKALIQILAAVAERDVVY 892

QY 901 FTFGDSELMRDIYSMTFLTERKLTVGE 928

Db 893 FTFGDSELMRDIYSMTFLTERKLDVGE 920

RESULT 9

O46043

ID O46043 PRELIMINARY; PRT; 768 AA.

AC O46043;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE PARG protein.

GN PARG OR EG:114E2.1 OR CG2864.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe C., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Ame J.-C., Jacobson M.K.;

RT "Isolation and characterization of the cDNA encoding Drosophila

RT poly(ADP-ribose) glycohydrolase.";

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 46-768 FROM N.A.

RA Murphy L., Harris D., Barrell B.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 46-768 FROM N.A.

RA Benos P.;

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003428; AAF45886.1; -

DR EMBL; AF079556; AAC28734.1; -

DR EMBL; Z98254; CAB10913.1; -

DR FlyBase; FBgn023216; Parg.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

KW Hydrolase.

SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;

Query Match 20.5%; Score 1063.5; DB 5; Length 768;

Best Local Similarity 41.5%; Pred. No. 6e-68;

Matches 244; Conservative 93; Mismatches 198; Indels 53; Gaps 15;

QY 417 RVPKAEKRRKEQCEMKHQRTKIPKYIPPHL--SPDKKWLGTPIEM--RRMRCGIRL 472

Db 45 RMSKSPDGGGISEIETEE-----PENLANSLLDSSWRGVSMIAHRNRQPFLENL 94

QY 473 PPLRPSANHTVTIRVDLLRIGCEVPKPPPTHFKDLNDKHKVMPCEONLYPVEDENGERA 532

Db 95 PPVTAGNLHRVMYQ---LPIRETP-PRPKSPGKWDSEHVRLLPCAPESKYPRENPDGSTT 150

QY 533 AGSRWELIQTALLNRLTRPQNLKDAILLKYNVAYSKKWDTALIDFWDKVLSEAEQAHLQ 592

Db 151 IDFRWEMIERALLQPIKTCCELQAALISYNTTYRDQWHFRALHQLLDELDESETRVFFE 210

QY 593 SILPDMVKIALCLPNICTQPIPLLKQKNHSITMSQEQIASLLANAFCTFPPRRNA-KMK 651

Db 211 DLLPRIRLAALRLPDLIQSPVPLKHKHKNASLSLSQQQISCLLANAFCLCTFPPRNTLKRK 270







Db 139 VFLSQELIGALLACSFCLFPDDNRGAK-----HLPVINFDHFLFASLYISYSQESKIRC 194  
QY 680 LFCYFRRVTEKKPTGLVTFTRQ--SLEDFPEWERCEKLLTRLHLVTVYEGTIEGNGQGMLOV 737  
Db 195 IMHYFERFCSCVPIGIVSFERKITAAPDADFWSKSDVSLCAFKVHSFGLIEDQPDNALEV 254  
QY 738 DFANRFVGGVTSAGLVQEEIRFLINPELIVSRFLTEVLDHNECLLIITGTQYSEYTGYA 797  
Db 255 DFANKYLGGSLSRGCVQEEIRFMINPELIIAGMLFLPRMDDNEAIEIVGAERFSCYTGYA 314  
QY 798 ETYRWARSHEDRSERDDQRRRTTEIVAIDAL-----HPRRYLDQFVPEKIRRELNKAYC 851  
Db 315 SSFRFAGEYIDKXAMPFKRRRTTRIVAIDALCTPKMRHFKDIC-----LLREINKALC 367  
QY 852 GFLR----- 855  
Db 368 GFLNCSKAWEHQNIFMDEGDNEIQLVRNGRDSGLLRTETTASHRTPLNDVEMNREKPPANN 427  
QY 856 -----PGVSSENL--SAVATGNWCGGAFGGDARLKALIQILAAAVAERDVV-YFTFGD 905  
Db 428 LIRDFYVEGVNDHEDDGVATGNWCGVFGGDPPELKATIQWLAASQTRRRPFISYYTFG- 486  
QY 906 SELMRDIYSMHTFLTERKLTGGEVYKLLRY 936  
Db 487 VEALRNLDQVTKWILSHKWTVDLWNMMLEY 517  
RESULT 12  
Q8VYA1 PRELIMINARY; PRT; 522 AA.  
AC Q8VYA1;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, last sequence update)  
DT 01-JUN-2003 (TremBLrel. 24, last annotation update)  
DE Putative poly(ADP-ribose) glycohydrolase.  
GN ATG331870.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopses.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
Chauk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
Ecker J., Theologis A., Davis R.W.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY072330; AAL61937.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;  
Query Match 10.3%; Score 533; DB 10; Length 522;  
Best Local Similarity 30.5%; Pred. No. 6.6e-30;  
Matches 132; Conservative 83; Mismatches 138; Indels 80; Gaps 11;  
QY 577 FWDKYLEEBAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKNH----- 622  
Db 80 FFDKKISREESANFFGEVVPALCRLLQLPSMLEXHY----QKADHVLDGVKSGRLRLGP 135  
QY 623 ----SITMSQEQIASLLANAFCTFPRRNAKMKSEYSSYPDINFNRLF-----GRSSRXP 674  
Db 136 QEAGIVLLSQELIAALLACSFCLFPFVDRSLK---NLQGINFSLGFSFPYMRHCTKQE 191  
QY 675 EKLKTLFCYFRRVTEKKPTGLVTFTRQSL-----DFPE---WERCEKLLTRLHVTY 723  
Db 192 NKIKCLIHVFGICRMWPTGTVSFERKILPLEYHPFVSYPKADSWANSVTPLCSIEIHT 251

QY 724 EGTIEGNGQGMLOVDPANRFVGGVTSAGLVQEEIRFLINPELIVSRFLTEVLDHNECLI 783  
Db 252 SGAIEDQPCBALEVDFADEYFGGLTSLSYDTLQEEIRFVINPELIIAGMIFLPRMDANEAI 311  
QY 784 ITGTEQYSEYTGAEYTYRWARSHEDRSERDDQRRRTTEIVAIDALHFRYRLDQFVPEKIR 843  
Db 312 IVGVERFSGYTGYPSPFQYAGDYTDNKLDFRRRKTRVIAIDAMP-DPGMGQYKLDALI 370  
QY 844 RELNKAYCGFLR-----PGVSS-----ENLSAVA 867  
Db 371 REVNKAFSGYMHQCKYNIDVKHDPEASSSHVPLTSDSASQVIESSHRWCIDHEEKIGVA 430  
QY 868 TGNWCGGAFGGDARLKALIQILAAAVAERDVV-YFTFGDSELMRDIYSMHTFLTERKLT 926  
Db 431 TGNWCGVFGGDPPELKIMLQWLAIQSQRPFMSYYTFG-LQALQNLNQVIEVMVALQEMTV 489  
QY 927 GEVYKLLRLRYNE 939  
Db 490 GDLWKKLVEYSSE 502  
RESULT 13  
Q867X0 PRELIMINARY; PRT; 781 AA.  
AC Q867X0;  
DT 01-JUN-2003 (TremBLrel. 24, Created)  
DT 01-JUN-2003 (TremBLrel. 24, last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, last annotation update)  
DE Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3  
protein) (Corresponding sequence F20C5.1a).  
GN F20C5.1 OR PME-3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode  
Caenorhabditis elegans.";  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Matthews P.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY185493; AAO26316.1; -.  
DR EMBL; Z68161; CAD89735.1; -.  
DR WormPep; F20C5.1a; CE333775.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 781 AA; 89273 MW; 30447D06605CED0A CRC64;  
Query Match 9.3%; Score 481.5; DB 5; Length 781;  
Best Local Similarity 24.1%; Pred. No. 6.2e-26;  
Matches 210; Conservative 126; Mismatches 297; Indels 239; Gaps 37;  
QY 199 DTNSEESRDNQQLTHVKLANAKQTME---DEQGREARSHQKCGKACHPAEACAGCQOE 254  
Db 10 DPTQDEKDYEDYV-GVGFAHQVPTMKRRLTEHGNTTESKE-----DPEE-----PK 56  
QY 255 ETDVVSSEPLSDTGSSED-----VGTGLKNA-----NRLNRQESSLGN 291  
Db 57 SRDVFVSSQSSDESQDSAEINPEIAKEVSENCENLTETLKISNIESLDNVTSEHILDN 116  
QY 292 ---SPFEEKSEPESPMDVDNSKNKSCQDSEA-----DEETSPGFDEQED----- 332

|           |  |   |              |
|-----------|--|---|--------------|
| Db        | 117  | HKSTEPMEEDVNNKSNIDV--AINSDDEDELVEENKEMRDGEVQVQDQDLFADDQELIEY  | 174          |
| QY        | 333  | -----SSSAQTANKPSRFQPREADTELKRSSAKGGEIRLHFQFEGGESRA            | 378          |
| Db        | 175  | PGIMKDTTQDITDSEVETAQKMEMIEETEADS-----TFVGEDSK-                | 216          |
| QY        | 379  | GMNDVNAKPGST-----SSLNVECRNSKQHGRKDSKITDHFMRVPAEDKRKEQCEM      | 431          |
| Db        | 217  | -----NQSQSGTSDVDADSQINLATKTVR-----TSSSSFLSTVSTCEAPAKGRARM     | 265          |
| QY        | 432  | KHQRTERKIPKYPHLS--PDKKWLGTPIEMRRMPCGIRLPLRPSANHTVTIRVDL       | 489          |
| Db        | 266  | YQKELEKHVIAFTGNLTLPD-----LNKVDPRDNY-----                      | 297          |
| QY        | 490  | LRIGEVKPPFP--THFKDLWNKH---VKMPCSEQNLVPEDENGERAAGSRWELIQTALL   | 545          |
| Db        | 298  | -RYCTIPN-FPASQGLREDNRYGPKIVLPQWRREF---DSRGR-----              | 337          |
| QY        | 546  | NRLTRPQNLKDALIKYN---VAYSKKWDFTA-----LIDFWD---KVLSEAEQAHLVQ    | 592          |
| Db        | 338  | -----RDSYFYFKRLDGLYKCYKTTGYFMFVGLLHNMWFEFDPDITYKLPALEMY       | 388          |
| QY        | 593  | SILPDMV-----KIALCLPNICTQPIPLKQKMNHSITMSQEQIASLLANAF           | 639          |
| Db        | 389  | KEMSELVGREVLEKFAVARIAKTAEDILPERIYLVGDV-ESATLSHKQCAALVARMF     | 447          |
| QY        | 640  | FCTFPRRNAMKSEYSSYPDINFNRNLFEGRSRKPEKLTLCYFRRVTEKKPTGLVTF-     | 698          |
| Db        | 448  | F-----ARPDSPFS-----FCRILSSDKSICVEKLFELTYFDKMSMDPPDGAVSFR      | 494          |
| QY        | 699  | -TRQSLDFPEWERCEKLLTRLHVY--EGTIEGNGQGMQLQVDFANRFVGGVTSAGLVQ    | 755          |
| Db        | 495  | LTKMDKDTFNEEWKDKLRSLEPEVEFFDEMLIEDTAL-CTQVDFANEHLGGVNLHGVSQ   | 553          |
| QY        | 756  | BEIRFLINPELIVSRLEFTEVLDHNECLITITGEQYSEYTYGAETYRWA-----RSHEDRS | 810          |
| Db        | 554  | BEIRFLMCPMMVGMMLCEKMKQLEAISIVGAYVFSSYTYGHTLKWAELOPNHSRQNTN    | 613          |
| QY        | 811  | E-RDDWQRRTTEIVAIDALHFR--YLDQFVPEKIRRELKAYCYGFLRPGVSSENLS      | 864          |
| Db        | 614  | EPRDRFGRRLRVETIAIDAILFKGSKLDCQTEQLNKNANIIREMKKASIGFMSQGPFTNI- | 672          |
| QY        | 865  | AVATGNWCGAFGGDARLKALIQILAAVAERDVVYTFPGDSELMRDIYSMHTFLTERKL    | 924          |
| Db        | 673  | PIVTGWGCGAFNGDKPLKFLIIQVIAAGVADRLPHFCFSFGEPELAACKKIIERMKQKDV  | 732          |
| QY        | 925  | TVGEVYKLLRYNNEECRNCSTPGPDIKLYPF                               | 956          |
| Db        | 733  | TLGMLFSMI-----NNTGLPHKHFFFYVF                                 | 756          |
| RESULT 14 |  |   |              |
| ID        | Q19637   | PRELIMINARY;  | PRT; 764 AA. |
| AC        | Q19637;  |   |              |
| DT        | 01-NOV-1996 (TrEMBLrel. 01, Created)                                 |   |              |
| DT        | 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)                    |   |              |
| DT        | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)                  |   |              |
| DE        | PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose      |   |              |
| DE        | metabolism enzyme-3 short form).                                     |   |              |
| GN        | F20C5.1 OR PME-3.  |   |              |
| OS        | Caenorhabditis elegans.  |   |              |
| OC        | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; |   |              |
| OC        | Rhabditidae; Peloderinae; Caenorhabditis.                            |   |              |
| OX        | NCBI_TaxID=6239;   |   |              |
| RN        | [1]  |   |              |
| RP        | SEQUENCE FROM N.A.   |   |              |
| RA        | Matthews P.;   |   |              |
| RL        | Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.              |   |              |
| RN        | [2]  |   |              |
| RP        | SEQUENCE FROM N.A.   |   |              |
| RX        | MEDLINE=99069613; PubMed=9851916;                                    |   |              |

|   |   |   |     |
|---|---|---|-----|
| RA  | none;   |   |     |
| RT  | "Genome sequence of the nematode C.elegans: A platform for            |   |     |
| RT  | investigating biology."   |   |     |
| RL  | Science 282:2012-2018(1998).  |   |     |
| RN  | [3]   |   |     |
| RP  | SEQUENCE FROM N.A.  |   |     |
| RA  | Gagnon S.N., Hardy I., Desnoyers S.;                                  |   |     |
| RT  | "Characterization of poly(ADP-ribose) glycohydrolases in the nematode |   |     |
| RT  | Caenorhabditis elegans."  |   |     |
| RL  | Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.               |   |     |
| DR  | EMBL; Z68161; CAA92299.2; --  |   |     |
| DR  | EMBL; AY185494; AAO26317.1; --  |   |     |
| DR  | PIR; T21138; T21138.  |   |     |
| DR  | WormPep; F20C5.1b; CE32867.   |   |     |
| DR  | InterPro; IPR007724; PARG.  |   |     |
| DR  | Pfam; PF05028; PARG; 1.   |   |     |
| SQ  | SEQUENCE 764 AA; 87441 MW; B391703B1D0CB79A CRC64;                    |   |     |
| Query Match 9.2%; Score 475; DB 5; Length 764;                      |   |   |     |
| Best local Similarity 24.0%; Pred. No. 1.8e-25;                     |   |   |     |
| Matches 208; Conservative 120; Mismatches 294; Indels 244; Gaps 36; |   |   |     |
| QY  | 199   | DTNSEESRDNQFLTHVLANAKQWME---DEQGREARSHQCKGKACHPAEACAGCQOE     | 254 |
| Db  | 10  | DPVTQDEKDYDV-VGGEFAHQVPTMKRRKLTGHTNTTESKE-----DPEE-----PK     | 56  |
| QY  | 255   | ETDVVSEPLSDTGED-----VGTGLKNA-----NRLNROESSLGN                 | 291 |
| Db  | 57  | SRDVFVSSQSSDESQEDSAENPEIAKEVSENCENTLTETLKISNIESLDNVTERSEHTLDN | 116 |
| QY  | 292   | ---SPPEKESEPESPMDV-----DNSKNSQDSEADEETSPGF--DEQE---331        |     |
| Db  | 117   | HKSTEPMEEDVNNKSNIDVAINSDDEDELVEENKEMRDGEVQVQDQDLFADDQELIE     | 176 |
| QY  | 332   | -----DSSAQTANKPSRFQPREADTELKRSSAKGGEIRLHFQFEGGESR             | 377 |
| Db  | 177   | YPGIMKDTTQDITDSEVETAQKMEMIEETEADS-----TFVGEDSK                | 219 |
| QY  | 378   | AGMNDVNAKPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPAEDKRKEQCEMKHQTE     | 437 |
| Db  | 220   | A-----TKTVRTSSSSFLST-----VSTCEAPAKGRARMYQKELE                 | 254 |
| QY  | 438   | RKIPKYIPPHLS--PDKKWLGTPIEMRRMPCGIRLPLRPSANHTVTIRVDLLRIGEV     | 495 |
| Db  | 255   | KHVIAFTGNLTLPD-----LNKVDPRDNY-----RYCTI                       | 285 |
| QY  | 496   | KPPFP--THFKDLWNKH---VKMPCSEQNLVPEDENGERAAGSRWELIQTALLNRLTRP   | 551 |
| Db  | 286   | PN-FPASQGLREDNRYGPKIVLPQWRREF---DSRGR-----320                 |     |
| QY  | 552   | QNLKDALIKYN---VAYSKKWDFTA-----LIDFWD---KVLSEAEQAHLVQSIPLDM    | 598 |
| Db  | 321   | ---RDSYFYFKRLDGLYKCYKTTGYFMFVGLLHNMWFEFDPDITYKLPALEMYKEMSEL   | 377 |
| QY  | 599   | V-----KIALCLPNICTQPIPLKQKMNHSITMSQEQIASLLANAFCTFPR            | 645 |
| Db  | 378   | VGREVLEKFAVARIAKTAEDILPERIYLVGDV-ESATLSHKQCAALVARMFF-----431  |     |
| QY  | 646   | RNAKMKSEYSSYPDINFNRNLFEGRSRKPEKLTLCYFRRVTEKKPTGLVTF--TRQSL    | 703 |
| Db  | 432   | --ARPDSPFS-----FCRILSSDKSICVEKLFELTYFDKMSMDPPDGAVSFLTKMDK     | 483 |
| QY  | 704   | EDFPEWERCEKLLTRLHVY--EGTIEGNGQGMQLQVDFANRFVGGVTSAGLVQEEIRFL   | 761 |
| Db  | 484   | DTFNEEWKDKLRSLEPEVEFFDEMLIEDTAL-CTQVDFANEHLGGVNLHGVSQEEIRFL   | 542 |
| QY  | 762   | INPELIVSRLEFTEVLDHNECLITITGEQYSEYTYGAETYRWA-----RSHEDRSE-RDDW | 815 |
| Db  | 543   | MCPMMVGMMLCEKMKQLEAISIVGAYVFSSYTYGHTLKWAELOPNHSRQNTNEFRDRF    | 602 |
| QY  | 816   | QRRTEIVAIDALHFR-----YLDQFVPEKIRRELKAYCGFLRPGVSSENLSAVATGN     | 870 |
| Db  | 603   | GRLRVETIAIDAILFKGSKLDCQTEQLNKNANIIREMKKASIGFMSQGPFTNI-PIVTGW  | 661 |

QY 871 WCGAAGGDAKALKALIQILAAVAERDVVYFTFGDSELMRDIVSMHTFLTERKLTGVEVY 930  
Db 662 WCGAAGGDAKALKALIQILAAVAERDVVYFTFGDSELMRDIVSMHTFLTERKLTGVEVY 930  
QY 931 KLLRYYNEECNCSTPGPDIKLYPF 956  
Db 722 SMI-----NNTGLPHKHFYVF 739

RESULT 15  
Q9N5L4 PRELIMINARY; PRT; 485 AA.  
AC Q9N5L4; PRELIMINARY; PRT; 485 AA.  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme  
4).

GN H23L24.5 OR PME-4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).

[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Abu-Threideh J., Lehnert L.;  
RL "The sequence of C. elegans cosmid H23L24.";  
RN [3]  
RP Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.

[4]  
RP SEQUENCE FROM N.A.  
RC Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Molecular characterization of poly ADP-ribose glycohydrolases in  
Caenorhabditis elegans.";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006662; AAF39896.2; -.  
DR EMBL; AF548468; AAN40699.1; -.  
DR WormPep; H23L24.5; CE32685.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;

Query Match 9.1%; Score 472; DB 5; Length 485;  
Best Local Similarity 29.9%; Pred. NO. 1.5e-25;  
Matches 143; Conservative 83; Mismatches 165; Indels 88; Gaps 17;

QY 537 WELIQTALLNRLTR-PQNLKDAALK--YNNVAYS-----KKWDTALIDF 577  
Db 46 FELLETGVSQQWNCQDLNFNEYLKTYKNGYSQFEDLLFKINGYSEKERFDLPALKSF 105  
QY 578 WDKVLE---EAEACHLYQSILPDMVKIALCLPNICTQPIPLKQKMN-----HSITMSQ 628  
Db 106 YRKMSEIVGEDE-----VLEKLARLVRTKSAC---EVLPEKIYRLVGDIESATFSH 154  
QY 629 EQIASLLANAFCTFPRRNAMKSEYSSYPDINPRLFEGRSSRKPEKLTLCYFRRVT 688  
Db 155 IQCASLIAMWFFSDTPR-----LSFIILQKTTCAVEKLFLETFYFDKMS 200

QY 689 EKKPTGLVTF--RQSEDFEWE-RCEKLLTLRLHVTYEGTIEGNGQMLQVDFA 741  
Db 201 IDPPIGAVSFRKMRITHKQYLEN---WKLRNTNLLPDVQVFDKMSIETAL-CTQIDFAN 256  
QY 742 RFVGGVTSAGLVQEEIRFLINPELIVSRLEFTEVLDHNECLIIITGTEQYSEYTGAEYR 801  
Db 257 KRLGGVLLKGGAVQEEIRFMCPMMVAAILLNDVTQDLEAISIVGAYVFSSYTGYSNTLK 316  
QY 802 WAR-----SHEDRSERDDMQRRITTEIVAIDALH-----FRYLDQFVPEKIRRELKAY 850  
Db 317 WAKITPKHSAQNNSFRDQFGRLOQTETVAIDAVRNAGTPLECLNQLTTEKLTREVRKAA 376  
QY 851 CGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMR 910  
Db 377 IGFLSAGDGFSKI-PVVGWGWCGGAFRGNKPLKFLIQVIACGISDRPLQFCTFGDTELAK 435  
QY 911 DIYSMHTFLTERKLTGVEVYKLLRLRYNEECRNCSTPGPDIKL-YPFYHAVESCTQTT 968  
Db 436 KCEEMTTLFRNNNVRTGQLFLII-----NSIGPPLNYSEQYVFDRAKINST 483

Search completed: May 26, 2004, 18:45:55  
Job time : 100.971 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:28:37 ; Search time 173.315 Seconds  
(without alignments)  
13270.296 Million cell updates/sec

Title: US-09-302-812-1  
Perfect score: 7242  
Sequence: 1 accggaagtgaaagcc.....aaatttcatttaacaaaaa 4070

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2 1/USPTO spool\_p/6333148/runat 26052004 150052 5654/app\_query.fasta\_1.12437  
-DB=A\_Geneseq\_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148.cgc 1\_1\_414 @runat 26052004 150052 5654 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description        |
|------------|--------|-------------|--------|----|--------------------|
| 1          | 5184   | 71.6        | 977    | 5  | Aae25629 Bovine po |
| 2          | 5184   | 71.6        | 977    | 5  | Aau76020 Bovine po |
| 3          | 5184   | 71.6        | 977    | 5  | Aau75799 Bovine po |
| 4          | 5184   | 71.6        | 977    | 6  | Abg72278 Bovine po |
| 5          | 4635.5 | 64.0        | 976    | 5  | Aae25630 Human pol |
| 6          | 4635.5 | 64.0        | 976    | 5  | Aau76021 Human pol |
| 7          | 4635.5 | 64.0        | 976    | 5  | Aau76012 Human pol |
| 8          | 4635.5 | 64.0        | 976    | 6  | Abg72279 Human pol |
| 9          | 4332.5 | 59.8        | 968    | 5  | Aae25631 Murine po |
| 10         | 4332.5 | 59.8        | 968    | 5  | Aau76022 Mouse pol |

|    |        |      |      |   |          |           |           |
|----|--------|------|------|---|----------|-----------|-----------|
| 11 | 4332.5 | 59.8 | 968  | 5 | AAU76013 | Aau76013  | Mouse pol |
| 12 | 4332.5 | 59.8 | 968  | 6 | ABG72280 | Abg72280  | Mmurine p |
| 13 | 1063.5 | 14.7 | 768  | 4 | ABB59491 | Abb59491  | Drosophil |
| 14 | 1063.5 | 14.7 | 768  | 5 | AAE25632 | Aae25632  | Fruit fly |
| 15 | 1063.5 | 14.7 | 768  | 5 | AAU76023 | Aau76023  | Fruit fly |
| 16 | 1063.5 | 14.7 | 768  | 5 | AAU76014 | Aau76014  | Fruit fly |
| 17 | 1063.5 | 14.7 | 768  | 6 | ABG72281 | Abg72281  | Fruit fly |
| 18 | 465    | 6.4  | 726  | 5 | AAE25633 | Aae25633  | Poly aden |
| 19 | 465    | 6.4  | 726  | 5 | AAU76024 | Aau76024  | Worm poly |
| 20 | 465    | 6.4  | 726  | 5 | AAU76015 | Aau76015  | Worm poly |
| 21 | 465    | 6.4  | 726  | 6 | ABG72282 | Abg72282  | C. elegan |
| 22 | 335.5  | 4.6  | 819  | 4 | ABG20721 | Abg20721  | Novel hum |
| 23 | 201    | 2.8  | 100  | 4 | ABG11103 | Abg11103  | Novel hum |
| 24 | 185.5  | 2.6  | 954  | 4 | ABB65602 | Abb65602  | Drosophil |
| 25 | 184    | 2.5  | 2665 | 4 | AAM14533 | Aam14533  | Peptide # |
| 26 | 184    | 2.5  | 2665 | 4 | ABB33490 | Abb33490  | Peptide # |
| 27 | 184    | 2.5  | 2665 | 4 | AAM26950 | Aam26950  | Peptide # |
| 28 | 184    | 2.5  | 2665 | 4 | ABB28314 | Abb28314  | Human pep |
| 29 | 184    | 2.5  | 2665 | 4 | ABB18950 | Abb18950  | Protein # |
| 30 | 184    | 2.5  | 2665 | 4 | AAM66665 | Aam66665  | Human bon |
| 31 | 184    | 2.5  | 2665 | 4 | AAM54270 | Aam54270  | Human bra |
| 32 | 184    | 2.5  | 2665 | 4 | ABG48336 | Abg48336  | Human liv |
| 33 | 184    | 2.5  | 2665 | 4 | AAM02259 | Aam02259  | Peptide # |
| 34 | 184    | 2.5  | 2665 | 5 | ABG36319 | Abg36319  | Human pep |
| 35 | 184    | 2.5  | 3266 | 3 | AAB42491 | Aab42491  | Human ORF |
| 36 | 184    | 2.5  | 3371 | 6 | ABO07211 | Abo07211  | Human p53 |
| 37 | 184    | 2.5  | 3654 | 6 | ABR47592 | Abra47592 | Breast ca |
| 38 | 184    | 2.5  | 3654 | 6 | ABO53027 | Abo53027  | Human put |
| 39 | 180    | 2.5  | 2194 | 4 | AAM40114 | Aam40114  | Human pol |
| 40 | 178    | 2.5  | 33   | 5 | AAE25652 | Aae25652  | Bovine po |
| 41 | 176    | 2.4  | 1210 | 7 | ABR63880 | Abr63880  | Human AF4 |
| 42 | 175.5  | 2.4  | 1163 | 4 | AAU28028 | Aau28028  | Novel hum |
| 43 | 174.5  | 2.4  | 768  | 4 | AAM78686 | Aam78686  | Human pro |
| 44 | 174.5  | 2.4  | 962  | 4 | ABG04842 | Abg04842  | Novel hum |
| 45 | 174.5  | 2.4  | 1462 | 6 | ABP58346 | Abp58346  | Human cel |

ALIGNMENTS

RESULT 1

AAE25629

ID AAE25629 standard; protein; 977 AA.

XX AC AAE25629;

XX DT 04-NOV-2002 (first entry)

XX DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).

XX KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.

XX OS Bos taurus.

XX PN US6395543-B1.

PD 28-MAY-2002.

PF 23-FEB-2000; 2000US-00511507.

PR 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

PA (KENT ) UNIV KENTUCKY RES FOUND.

PI Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

DR N-PSDB; AAD42081.



XX

PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage.  
PT Inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX

PS Claim 3; Col 47-45; 77pp; English.

XX

CC The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The antisense therapeutic are useful to treat neoplastic disorders and conditions caused by genotoxic oxidative stress e.g., cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. The invention is useful in gene therapy and antisense therapy. The present sequence is bPARG

XX

SQ Sequence 977 AA;

Alignment Scores:

Pred. No.: 0 Length: 977  
Score: 5184.00 Matches: 977  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 71.58% Indels: 0  
DB: 5 Gaps: 0

US-09-302-812-1 (1-4070) x AAE25629 (1-977)

QY 258 ATGAGTGGGCGCCCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGACGCGCTGCA 317  
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
QY 318 ACTTCTCCGCGCGCCCTCGGACGCCCGGAGCTTCCCGGCGAGCGCGCTCCTC 377  
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
QY 378 GATTCCAGGACGCTCCGGTGCAGTTCAGGGTCCCGCGCTCCTCGTCAGGCTGCGCCTG 437  
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
QY 438 GGCCGGCGGACAGCACGAGGCGAGGCGCCACCTCTCTTGTGTTTCAAACAGAGACTATA 497  
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
QY 498 ACCAGTGGATGGACACTAAAGGAATCAAGACAGTGTGAATCAGAAAGTTTGATAGTAAA 557  
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
QY 558 GAAACAAACAATACAGAGAGAGAAATCAGATGAGTCTCTGACAAAAGATAACTTTAT 617  
Db 101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120  
QY 618 CAACATAACATGGAAAATTAGAAAATGTTTCTCAGCTAGTTTGTGATAAGTCACAGTT 677  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 678 GAAAGAGTACACAGTATTTGAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737  
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 738 GAAGGCCACACTCAGAACCGGTTTTGAAAGTGAACCTCCAGCGGTAACTCTGTACCA 797  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 798 GAGCAGTTTCAATATGTTATGTCAGTCAGTCGTCCTCCCAAGGATGATCACAGTGACACA 857  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

QY 858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAAGCTTGGAGATGCA 917  
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAGCCAGAACCCAGAGAGTGTGGCAAG 977  
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 978 GCTTGCCATCCTGCAGAACCTGTGCAGGGTGTGCAGGAGGAGACAGACGCTGGTGTCC 1037  
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValSer 260  
QY 1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGATGTTGGTACTGGACTGAAAAATGCCAAC 1097  
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
QY 1098 AGATTGAATAGACAAAGAGTAGTCTAGGAAATTTCTCTCATTGTGAAAGAAAGTGAA 1157  
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCAAAAATAGTTGTGAGATTGAGAGCAGAT 1217  
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
QY 1218 GAAGAGACAAAGTCCAGGTTTGTGATGAACAGGAAGATAGCAGTTCTGCTCAACAGCAAAT 1277  
Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAAGCGGTCTCTGCT 1337  
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
QY 1338 AAGGAGGTGAGATTGATTACATTTCCAATTTGAAGGAGAGAGAGTTCGAGCTGGAATG 1397  
Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
QY 1398 AATGATGTGAATGCCAAACGAGCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTTCTAAAATCACAGATCATTTTCATGAGAGTGCCCAA 1517  
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
QY 1518 GCAGAGGACAAAAGAAAAGAACAAATGTGAATGAAACATCAAGATCATTTTCATGAGAGTGC 1577  
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
QY 1578 CTTAAATACATTCACCTCCACCTTTCTCCAGATAGAAATGGCTTGGAACTCCTATTGAG 1637  
Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
QY 1638 GAGATGAGGAGATGCCAAGGTGTGGGATCCGGCTCCCTCCCTTGAGACCATCTGCCAAT 1697  
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
QY 1698 CACACAGTACTATTCGGGTAGATCTTTTGGCAATAGGAGAGTTTCTAAACCTTTCCCA 1757  
Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProphePro 500  
QY 1758 ACACATTTTAAAGATTGTGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAC 1817  
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
QY 1818 TTGTACCTGTGGAAGATGAGAATGGTGAAGAGCTGCAGGAGCCCGTGGAACTCAT 1877  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
QY 1878 CAGACTGCATCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

QY 1938 TACAATGTGGCATATTTCTAAGAAATGGGACTTTACAGCTTTTGATTGATTTCTGGGATAAG 1997  
Db 561 TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuIleAspPheTrpAspLys 580  
QY 1998 GTACTAGAAGAAGCAGACGCTCAACACTTGATCAGTCCATCTTGCCCTGATATGCTGAAA 2057  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
QY 2058 ATTCGACTCTGTCTGCCAATAATTTGTACCCAGCCATACCACTCCTCGAAACAGAAATG 2117  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
QY 2118 AATCATTTCCATCACAATGTCAAGGAAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
QY 2178 TGCACGTTTCCACGACGCAATGCCAAGATGAATCAGAGTATTCAGATTATCCAGATATT 2237  
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
QY 2238 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACACAGAGAAGCTTAAACGCTC 2297  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
QY 2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAACCCACTGGTGGTGACATTTCACAAGA 2357  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACTCTGACTCGACTGCAT 2417  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
QY 2418 GTCACTTACGAAGGTACCATAGAGGAACCGCCAGGGCATGCTACAGGTGGATTTTGCA 2477  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2478 AACCGTTTCGTTGGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAGAAATCCGCTTT 2537  
Db 741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
QY 2538 TTAATCAACCTGAGTTGATTTTCCAGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2598 TGTCTTATCATCAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCGGAAACATAC 2657  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2658 CGCTGGGCGCGGACCATGAAGACAGGAGCGAAGGACGACTGGCAGAGCGCAGACT 2717  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgArgThrThr 820  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACAGTTTGTGCCCCGAG 2777  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTCTTCGTCCTGGAGTTCTTCA 2837  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
QY 2898 AGACTAAAGCCTTAATACAGATCTTGGCAGCTGCTGTAGTGCAGGAGACGTGGTGTAT 2957  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGACATTTACAGATGCATATCTCTCACT 3017  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 3018 GAGAGGAAACTGACTGTTGGAGAGTATATAAGCTGTCTGCTACGATATTACAATGAAGA 3077

Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuArgTyrTyrAsnGluGlu 940  
QY 3078 TGCAGAAACTGCTCCACCCCGGACGACGACATCAAGCTTTATCCATTATATACCATGCA 3137  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrPropheIleTyrHisAla 960  
QY 3138 GTTGAGTCTGTACACAGACCACCAACGCGCGGACAAAGGACGGGGCC 3188  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977  
RESULT 2  
AAU76020  
ID AAU76020 standard; protein; 977 AA.  
XX AC AAU76020;  
XX 08-MAY-2002 (first entry)  
DT XX  
DE Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX Bos taurus.  
OS  
XX  
FH  
FT Key Location/Qualifiers  
Region 601..617 /note= "Represents PARG oligopeptide #1"  
Region 761..770 /note= "Represents PARG oligopeptide #2"  
Region 771..801 /note= "Represents PARG oligopeptide #3"  
Region 849..880 /note= "Represents PARG oligopeptide #4"  
XX US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-163240/21.  
DR N-PSDB; ABK14931.  
XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX Claim 2; Col 47-52; 81pp; English.  
PS  
XX The present invention relates to a new poly (ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.









QY 318 ACTTCTCCGCGCGCCTCGGACGCCCGAGCTTCCCGGAGGCGGAGCGGCGCTCCTC 377  
Db 21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
QY 378 GATTCCAAAGGAGCGCTCCGGTGCAGTTCAGGTCGCCCGCTCCTCGTCAGGCTGCGCCCTG 437  
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
QY 438 GGCCGGCGGACAGCAGCACCGGAGGCGGCCACTCTCTTGTGTTTCAAAAGAGACTATA 497  
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTCAGTTCAGAAAGTTGCATAGTAA 557  
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
QY 558 GAAACAACAATAAGAGAAATCCATGATGAGTTCGTACAAAAGATAACTTTTAT 617  
Db 101 GluAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120  
QY 618 CAACATAACATGGAATAATTAGAAAATGTTTCTCAGCTAGGTTTGTGATAAGTCACCAGTT 677  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 678 GAAAAAGGTACACATATTGAAGCAGCATCAGACTCGGCTATGTGTAGTGGCAGAAT 737  
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGTTACCA 797  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCAGTCGCCCAAAGGATGATCAGAGTCACACA 857  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
QY 858 AATAGTGAGGAGAGTAGAGATAATCAGCAGCTTTTTCACACATGTAAAGCTTCGGAATGCA 917  
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 918 AAGCAGACGATGGAAAGATGAACAGGGCAGAGAGCCAGAGAGCCACCAGAAAGTGTGGCAAG 977  
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 978 GCTTGCCATCCTGTCAGAACCTGTGTCAGGCTGTACAGGAGGAGACAGACGCTGGTGTC 1037  
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260  
QY 1038 GAGAGCCCTTGTCCGACACTGGCTCTGAGGATGTGGTACTGGACTGAAATAATGCCAAC 1097  
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCCTCCATTGAGAAAGAAAGTGAA 1157  
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
QY 1158 CCTGAGTCACCAATGGATGTAGATAAATCCAAAAATAGTTGTGAGGATTCAGGATTCAGAGCAGAT 1217  
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
QY 1218 GAAGACACAAGTCCAGGTTTTCATGAACAGGAAGATAGCAGTTCGTCAAAACAGCAAAAT 1277  
Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAAGCGGTCCTTGCT 1337  
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
QY 1338 AAGGAGGTGAGATTTCGATTACATTTCCAAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397  
Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457

Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTTAAAAATCACAGATCAATTCATGAGAGTGCCCAAA 1517  
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
QY 1518 GCAGAGGACAAAAAGAAACAATGTGAAATGAAACATCAAAAGAACAGAAAGGAAGATC 1577  
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
QY 1578 CCTAAATACATTCCACCTCACCTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637  
Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
QY 1638 GAGATGAGGAGAATGCCAAGGTGTGGATCCGGCTGCCCTCCCTTGAGACCATCTGCCAAT 1697  
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
QY 1698 CACACAGTGACTATTTCGGGTAGATCTTTTGGCGAATAGGAGAAAGTTCCTAAACCTTTCCCA 1757  
Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
QY 1758 ACACATTTTAAAGATTGTGGGACACAAGCATGTTAAGATGCCTTGTTCAGAAACAAAAC 1817  
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
QY 1818 TTGTACCTCTGGAAGATGAGAATGGTGAGCGAGCTGCAGGAGCCGGTGGGAACTCATT 1877  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
QY 1878 CAGACTGCACCTTCTCAACAGGCTCAGTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTTGATTGATTCTGGGATAAG 1997  
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
QY 1998 GTACTAGAAAGACAGAAAGCTCAACACTGTATCATCAGTCCATCTTTCCTGATATGGTAAA 2057  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
QY 2058 ATTGCACCTCTGTCTGCCAAATATTTGTATCCAGCCCAATACCACCTCCTGAAACAGAAATG 2117  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
QY 2118 AATCATTCATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTTAATGCTTTCTTC 2177  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
QY 2178 TGCACGTTTCCACAGCGCAATGCCAAGATGAAATCAGAGTATTCAGTATCCAGATATT 2237  
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
QY 2238 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACCCACTGGGTGGTGACATTCACAAAG 2297  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
QY 2298 TTCTGTACTTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTGGTGACATTCACAAAGA 2357  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2358 CAGAGTCTTGAAGATTTCCAGAGTGGGAAAGATGTGAAAACTCCTGACTCGACTGCAT 2417  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
QY 2418 GTCACTTACGAAGGTACCATAGAGGAAACGGCCAGGCGCATGCTACAGGTGGATTTTGCA 2477  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2478 AACCGTTTCGTTGGAGGTGGTGTAAACCAGTGCAGGACTTGTGCAAGAGAAATCCGCTTT 2537

Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluLeuArgPhe 760  
QY 2538 TTAATCAACCTGAGTTGATTGTTTACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2598 TGTCTTATCATACAGGTACTGAGCAGTACAGTACAGTACAGGTATGTCGCAACATAC 2657  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2658 CGCTGGCCCGGAGCCATGAAGACAGGAGCGAAGGACGACTGGCAGGCGCACGACT 2717  
Db 801 ArgTyrAlaArgSerHisGluAspArgSerGluArgAspTyrGlnArgArgThrThr 820  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCACTTTGTGCCCGAG 2777  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGATTTCTTCTGCTCGTGGAGTTCTTCA 2837  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2838 GAGAACCTGTCTGAGTGGCTACAGGAACTGGGCTGTGGTGCCTTTGGGGTGATGCT 2897  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
QY 2898 AGACTAAAGCCTTAATACAGATCCTGGCAGCTGCTGAGTGGAGACGCTGTTTAT 2957  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATATCTCTCACT 3017  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 3018 GAGAGGAACTGACTGTGGAGAAGTATATAAGCTGCTGTACGATATACATGAAGAA 3077  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
QY 3078 TGCAGAAACTGTCTCCACCCCGGACCAGACATCAAGCTTTATCCATTATACATGCA 3137  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 3138 GTTGAGTCTGTACACAGACCAACCAACGCGGAGCAAGGACGGGGCC 3188  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

RESULT 4  
ABG72278

ID ABG72278 standard; protein; 977 AA.

XX XX  
AC ABG72278;

DT 13-MAR-2003 (first entry)

DE Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.

XX  
KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective; enzyme.

XX Bos taurus.

XX US2002132328-A1.

XX 19-SEP-2002.

XX 09-OCT-2001; 2001US-00973451.

XX

PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AMEJ J.  
PA (LINW/) LIN W.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2003-155895/15.  
XX N-PSDB; ABX14477.  
PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX Claim 28; Fig 16; 86pp; English.  
PS The present invention relates to the isolation of poly(ADP-ribose)  
XX glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents bovine PARG enzyme  
XX Sequence 977 AA;  
SQ

Alignment Scores:

Pred. No.: 0 Length: 977  
Score: 5184.00 Matches: 977  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 71.58% Indels: 0  
DB: 6 Gaps: 0

US-09-302-812-1 (1-4070) x ABG72278 (1-977)

QY 258 ATGAGTGGGGCCCGGCTGTGAGCCCTGCACCAAGCACCCCGCTGGAGCGCGCTGCA 317  
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
QY 318 ACTTCTCCGCGCGCTCGGACGCGCGAGCTTCCCGGCGAGGAGCGCGCTCTC 377  
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgVal 40  
QY 378 GATTCCAAGGACGCTCCGCTGAGTTCAGGGTCCCGGCTCTCGTCAGGCTGCGCCCTG 437  
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAla 60  
QY 438 GGCCGGGGGACAGCAGCAGCGGCGCCACCTCTTGTCTTCAACAGAACTATA 497  
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThr 80  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGATCAGAAAGTTTCATAGTAAA 557  
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
QY 558 GAAAAACAACAATACAGAGAGAATCCATGATGAGTCTGTACAAAAGATACTTTAT 617  
Db 101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

QY 618 CAACATAACATGGAAAAATTAGAAAAATGTTTCTCAGCTAGGTTTGTGATAAGTCACCAGTT 677  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 678 GAAAAAGGTACACAGTATTTTGAAGCAGCATCAGACTGCGGCTATGTGAAGTGGCAGAA 737  
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 738 GAAGGGCCCACTCAGAACCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCTCCCAAGGATGATCAGTGACACA 857  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
QY 858 AATAGTGAGGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAAGCTTGCGAATGCA 917  
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAACCCAGAACCCACCAGAGTGTGGCAAG 977  
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 978 GCTTGCCATCCTGCAGAACCTGTGCAGGGTGTGCAGGGAGGAGACAGACGCTGGTGTC 1037  
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
QY 1038 GAGAGCCCCTTGTCCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097  
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTGAGAAAGAAAGTGAA 1157  
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
QY 1158 CCTGAGTCACCAATGGATGATGATAAATCCAAAAATAGTTGTGAGGATTCAGAACGAGAT 1217  
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
QY 1218 GAAGAGACAAGTCCAGGTTTGTGATGAACAGGAAGATAGCAGTTCTGCTCAAAACAGCAAA 1277  
Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerAlaGlnThrAlaAsn 340  
QY 1278 AAACCTTCAAGTTCACCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCTCTGCT 1337  
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
QY 1338 AAGGAGGTGAGATTGATTACATTCCAAATTTGAAGGAGGAGAGAGTGCAGCTGGAATG 1397  
Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAAGAA 1457  
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTAAATCACAGATCATTTTCATGAGAGTGCCCAAA 1517  
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
QY 1518 GCAGAGGACAAAAGAAAAGAACAAATGTGAATGAACATCAAGAACAGAAAGGAAGATC 1577  
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
QY 1578 CCTAAATACATTCACCTCACCTTCTCCAGATAAGAAATGGCTTGGAACCTCCTATTGAG 1637  
Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGGATCCGCTGCTCCCTTGAGACCATCTGCCAAT 1697  
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
QY 1698 CACACAGTGACTATTTCGGGTAGATCTTTTGGCAATAGGAGAAGTTCTTAAACCTTTCCCA 1757

Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
QY 1758 ACACATTTTAAAGATTGTGGACAAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAAC 1817  
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
QY 1818 TTGTACCTGTGGAAGATCAGAATGGTGAGCGAGCTGCAGGAGCCGGTGGAACTCAT 1877  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
QY 1878 CAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
QY 1938 TACAATGTGGCATATTCTTAAGAAATGGGACTTTACAGCTTTTGATTGATTCTGGGATAAG 1997  
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
QY 1998 GTACTAGAAGAACGAGAACGCTCAACTTGTATCAGTCCATCTTCCTGATATGGTAAA 2057  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
QY 2058 ATTGCACCTCTGTCTGCCAAATATTTGTACCCAGCCCAATACCACTCTCTGAAACAGAGATG 2117  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
QY 2118 AATCATTTCCATCACAATGTCCAGGAACAGATTCAGATATTCAGATATTCAGATATTCAGATAT 2177  
Db 621 AsnHisSerIleThrMetSerGlnGlnGluIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
QY 2178 TGCACGTTTTCCACGACGCAATGCCAAGATGAAATCAGAGATATTCAGATATTCAGATATTCAGATAT 2237  
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
QY 2238 AACTTCAATCGGTTGTTGAAGGACGTTTCATCAAGGAAACCAGAGAAAGCTTAAACCGCTC 2297  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
QY 2298 TTCTGCTACTTTAGAAAGTCAACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAAGA 2357  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2358 CAGAGTCTTGAAAGATTTCAGAGTGGGAAAGATGTGAAAACTCCTGACTCGACTGCAT 2417  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
QY 2418 GTCACTTACGAAGGTACCATAGAAGGAAACCGCCAGGCGATGCTACAGGTGGATTTTGA 2477  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2478 AACCGTTTCGTTGGAGTGGTGTAAACCAGTGCAGGACTTGTGCAAGAAATCCGCTTT 2537  
Db 741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
QY 2538 TTAATCAACCCCTGAGTTGATTGTTTACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2598 TGCTTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATCCCGAAACATAC 2657  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2658 CGCTGGGCCCGAGCCATGAACAGAGGCGAAAGGAGCGACTGGCAGAGGCGCAGACT 2717  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCGAGTTTGTGCCCGAG 2777  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2778 AAGATCAGACGGGAGCTTAAACAGGCTTACTGTGGATTCTTCGTCTCGAGTTTCTTCA 2837



Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2838 GAGAACCTGTCTGAGTGGCTACAGGAACTGGGCTGTGGTCTTGGGGTGATGCT 2897  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
QY 2898 AGACTAAAGCCTTAATACAGATCCTGGCAGCTGTAGCTGAGCGAGACGTGTTTAT 2957  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValTyr 900  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 3018 GAGAGGAACTGACTGTGGAGAGATATATAAGCTGTCTGCTACGATATTACAATGAAGAA 3077  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
QY 3078 TGCAGAAACTGCTCCACCCCGGACCAGACATCAAGCTTTATCCATTTCATATACCATGCA 3137  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 3138 GTTGAGTCTGTACACAGACCAACCAACGCGGACAAAGGACGGGGCC 3188  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

RESULT 5  
AAE25630  
ID AAE25630 standard; protein; 976 AA.  
XX AC AAE25630;  
XX DT 04-NOV-2002 (first entry)  
XX DE Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
XX KW Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
KW cytosolic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX OS Homo sapiens.  
XX PN US6395543-B1.  
XX PD 28-MAY-2002.  
XX PF 23-FEB-2000; 2000US-00511507.  
XX PR 01-MAY-1998; 98US-0083768P.  
XX PR 30-APR-1999; 99US-00302812.  
XX PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
XX N-PSDB; AAD42082.  
XX PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX Claim 3; Col 55-60; 77pp; English.  
XX PS The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.

CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is human PARG  
XX Sequence 976 AA;  
SQ

Alignment Scores: Length: 976  
Pred. No.: 0 Matches: 874  
Score: 4635.50  
Percent Similarity: 92.83% Conservative: 32  
Best Local Similarity: 89.55% Mismatches: 69  
Query Match: 64.01% Indels: 1  
DB: 5 Gaps: 1

US-09-302-812-1 (1-4070) x AAE25630 (1-976)  
QY 258 ATGAGTGGGGCCCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCGCTGCA 317  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
QY 318 ACTTCTCCCGCCCGCTCGGACGCCCGGAGTTCCTCCCGGACGAGCGCGCTCCTC 377  
Db 21 ThrSer--ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39  
QY 378 GATTCCAAGGACGCTCCGGTGCAGTTTCAGGGTCCCGCTCCTCGTCAGGCTGCGCCCTG 437  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 438 GGCCGGGGGACAGCACCGGACCGGACGCCACCTCTCTTGTTCCTCAAAACAGAAGACTATA 497  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTCATAGTAAA 557  
Db 80 ThrSerTrpMetAspThrLysGlyLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 558 GAAACAAACAATACAGAGAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTTAT 617  
Db 100 GluAsnAsnAsnThrArgIleGluSerMetSerSerValGlnLysAspAsnPheTyr 119  
QY 618 CAACATAACATGGAATAATAGAAATGTTTCTCAGTAGTTTGTGATAAGTCACAGTT 677  
Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 678 GAAAGAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAACTGGCAGAAT 737  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 738 GAAGGGCCACTCAGAACCGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCTCCTCCCAAGGATGATCAGTGCACACA 857  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 858 AATAGTGAGGAGTAGAGATAATCAGCAGTTTTCACACATGTAAAGCTTGGCAATGCA 917  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAGCCAGAACCCAGAGAGTGTGGCAAG 977  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 978 GCTTCCCATCCTGCAGAGAGCTGTGAGGGTGTGAGAGGAGGACAGACGTTGGTGTC 1037  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluLeuAspValPro 259



QY 1038 GAGAGCCCTTGTTCGGACACTGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097  
:::|||||  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 1098 AGATTGAATAGACAAGAAAGTAGCTAGGAAATTCTCTCCATTTGAGAAAGAAAGTGAA 1157  
:::|||||  
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
QY 1158 CCTGAGTCAACCAATGGATGTAGATAATCCAAAAATAGTTGTTCAGGATTCAGAAGCAGAT 1217  
|||||  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 1218 GAAGAGACAAGTCCAGGTTTGTATGAACAGGAAGATAGCAGTTCTGCTCAACAGCAAAAT 1277  
|||||  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
QY 1278 AAACCTTCAAGTTCCAACCAAGAGAGAGCTGACACTGAGTTGAGGAAGCGGTCTCTGCT 1337  
|||||  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1338 AAGGAGGTGAGATTCCGATTACATTTCCAAATTTGAAGGAGGAGAGAGTCCGAGCTGGAATG 1397  
|||||  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
|||||  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAAATCAGACATCATTTTCATGAGAGTGCACAAA 1517  
|||||  
Db 400 SerLysGlnHisGlyLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1518 GCAGAGGACAAAAGAAAGAAACAATGTGAATGAACATCAAGAACAGAAAGAGATC 1577  
|||||  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1578 CCTAAATACATTCACCTCACCTTCTCCAGATAAGAAATGGCTTGGAACCTCCTATTGAG 1637  
|||||  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1638 GAGATGAGGAGAATGCCAAGGTGTGGATCCGGCTGCCCTCCCTTGAGACCATCTGCCAAT 1697  
|||||  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1698 CACACAGTGAATATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCTAAACCTTTCCCA 1757  
|||||  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1758 ACACATTTAAAGATTGTGGGACACAAACAGCATGTTAAGATGCCTTGTTCAGAAACAAAC 1817  
|||||  
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1818 TTGTACCTGTGGAAGATGAGAATGGTGACGAGCTGCAGGCGCGGTGGGAACCTCATT 1877  
|||||  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1878 CAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
|||||  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTTGATTGATTCTGGGATAAG 1997  
|||||  
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1998 GTACTAGAAGAAGCAGAAAGCTCAACACTTGTATCAGTCCCATCTTGCTGATATGGTGAAA 2057  
|||||  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 2058 ATTGCACCTCTGTCTGCCAAATATTGTACCCAGCCCAATACCACCTCTCTGAAACAGAAAGATG 2117  
|||||  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 2118 AATCATTCATCACAAATGTTCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177

Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGTTATCCAGATATT 2237  
Db 640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 2238 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAACACAGAGAAGCTTAAACCGCTC 2297  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2298 TTCTGCTACTTTAGAAAGAGTCACAGAGAAAAAACCCACTGGGTGGTGACATTCACAAGA 2357  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2358 CAGAGTCTTGAAGATTTCAGAGTGGGAAAGATGTGAAAAAACTCCTGACTCGACTGCAT 2417  
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2418 GTCACCTACGAAGGTACCATAGAAAGAAACGGCCAGGCGCATGTACAGGTGGATTTTGCA 2477  
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2478 AACCGTTTCGTTGGAGGTGTGTAAACCAGTGCAGGACTTGTGCAAGAAAGAAATCCGCTTT 2537  
Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
QY 2538 TTAATCAACCCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db 760 LeuIleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTACAGTGAATACACAGGCTATCCGAAACATAC 2657  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2658 CGCTGGGCCCGGACCCATGAAGACAGGACAGGAGCGAAAGGACGACTGGCAGAGCGCACGACT 2717  
Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2778 AAGATCAGACGGGAGCTTAAACAGGCTTACTGTGATTTCTTCGTCCIGGAGTTCTTCA 2837  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2838 GAGAACCTGTCTGCAGTGGCTACAGAAACTGGGGCTGTGTGCTTGGGGGTGATGCT 2897  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2898 AGACTAAAAAGCCTTAATACAGATCCTGGCAGCTGTGTAGCTGAGCGAGACGTGGTTTAT 2957  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaArgAspValValTyr 899  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCTCCTCACT 3017  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 3018 GAGAGGAAACTGACTGTTGGAGAGTATATATAAGCTGTCTACGATATTACAATGAAGAA 3077  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 3078 TGCAGAAACTGCTCCACCCCGGACACAGACATCAAGCTTTATCCATTATATACCATGCA 3137  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 3138 GTTGAGTCTGTACACAGACCACCAACCGCGGACAAAGACGGGG 3185  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975  
RESULT 6  
AAU76021

ID AAU76021 standard; protein; 976 AA.  
XX  
AC AAU76021;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
DE  
XX Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
OS Homo sapiens.  
XX  
PN US6337202-B1.  
XX  
XX 08-JAN-2002..  
XX  
XX 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI  
XX WPI; 2002-163240/21.  
DR  
DR N-PSDB; ABK14932.  
XX  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
XX Claim 2; Col 55-60; 81pp; English.  
XX  
XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the human PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX  
SQ Sequence 976 AA;

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4635.50 Matches: 874  
Percent Similarity: 92.83% Conservative: 32  
Best Local Similarity: 89.55% Mismatches: 69  
Query Match: 64.01% Indels: 1  
DB: 5 Gaps: 1

US-09-302-812-1 (1-4070) x AAU76021 (1-976)

QY 258 ATGAGTGGGCGCCCGGTGTGAGCCCTGCACCAAGCGACCCCGTGGGACCGCGTGA 317  
|||::|||  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
|||::|||  
QY 318 ACTTCTCCGCGCGCCCTCGGACGCCCGGAGCTTCCCGGCGAGGCGGCGTCTC 377  
|||::|||

Db 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 378 GATTCCAAGGACGCTCCGGTGCAGTTTCAGGGTCCCGCCGCTCTCTCGTCAGGCTGCGCCCTG 437  
|||  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
|||  
QY 438 GGCCGGCGGGGACAGCACCCAGGCGAGCCACCTCTCTTTGTTTCAACAGAGACTATA 497  
|||::|||  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
|||  
QY 498 ACCAGTTGGATGGACACTAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557  
|||  
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
|||  
QY 558 GAAACCAACAATACAAGAGAAATCCATGATGAGTTCTGTACAAAAGATAAATTTTAT 617  
|||  
Db 100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
|||  
QY 618 CAACATAACATGGAATAATTAGAAATGTTTCTCAGCTAGGTTTTCATATAAGTCACCAGTT 677  
|||::|||  
Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
|||  
QY 678 GAAAGAGGTACACAGTATTGGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737  
|||  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
|||  
QY 738 GAAGGGCCCACTCAGAACCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
|||  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
|||  
QY 798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCTCCCAAGGATGATCAGTGCACACA 857  
|||  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
|||  
QY 858 AATAGTGAGGACAGTAGATAATCAGCAGTTTGTGACACATGTAAAGCTTCCGAATGCA 917  
::|||  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
|||  
QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAGCCAGAACCCAGAGTGTGGCAAG 977  
|||  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
|||  
QY 978 GCTTGCCATCCTGCAGAACCTGTGCAGGGTGTGCAGGAGGAGGAGACAGACGCTGTGTC 1037  
::|||  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
|||  
QY 1038 GAGAGCCCTTGTGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAATAATGCCAAC 1097  
::|||  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
|||  
QY 1098 AGATTGAATAGACAAAGAGTAGTCTAGGAAATTCCTCCATTTGAGAAAGAAAGTGAA 1157  
::|||  
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPropheGluLysGluSerGlu 299  
|||  
QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAATAATAGTTGTCAGGATTCAGAGCAGAT 1217  
|||  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
|||  
QY 1218 GAAGAGACAAGTCCAGGTTTGTATGAACAGGAGATAGCAGTTCTGCTCAACAGCAAT 1277  
|||  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
|||  
QY 1278 AAACCTTCAAGGTTCCACCAAGAGAGCTGACACTGAGTTGAGGAAGCGGTCTCTGCT 1337  
|||  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
|||  
QY 1338 AAGGAGGTGAGATTGATTACATTTCCAATTTGAAGGAGGAGAGAGTTCGAGCTGGAATG 1397  
|||  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGluSerArgThrGlyMet 379  
|||  
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
|||  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerSerLeuAsnValGluCysArgAsn 399  
|||

QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTAAATCACAGATCATTTTCATGAGAGTGCCCAAA 1517  
Db SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1518 GCAGAGGACAAAAGAAAAGAAATGTGAATGAAACATCAAGAAACAGAAAGGAAGATC 1577  
Db AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1578 CCTAATACATTCACCTCCACTTCTCCAGATAAGAAATGGCTTGAACTCCTATTGAG 1637  
Db ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGGGATCCGGTCCCTCCCTTGAGACCATCTGCCAAT 1697  
Db GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1698 CACACAGTGACTATTCCGGGTAGATCTTTTGGCAATAGGAGAGTTCTTAAACCTTCCCA 1757  
Db HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1758 ACACATTTTAAAGATTTGTGGGACAACACAGCATGTTAAGATGCCCTTGTTCAGAACAAAC 1817  
Db ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1818 TTGTACCTGTGGAAGATGAGAAATGGTACCGGAGCTGCAGGCAGCCGGTGGAACTCATT 1877  
Db LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1878 CAGACTGCATCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
Db GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTCTGGGATAAG 1997  
Db TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1998 GTACTAGAAGAACGACGAGCTCAACACTTGATCAGTCCATCTTGCTGATATGGTGAAA 2057  
Db ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 2058 ATTGCACCTGTCTGCCAAATATTTGTACCCAGCCCAATACCCTCCTGAAACAGAAAGATG 2117  
Db IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 2118 AATCATTCATCACAAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177  
Db AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGTTATCCAGATATT 2237  
Db CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 2238 AACTTCAATCGGTTGTTTGAAGACGTTTCATCAAGGAAACAGAGAAGCTTAAACGCTC 2297  
Db AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2298 TTCTGTACTTTAGAGAGTTCACAGAGAAAACCCACTGGGTGGTGACATTCACAAGA 2357  
Db PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACCTCCTGACTCGACTGCAT 2417  
Db GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2418 GTCACTTACGAAGGTACCATAGAGGAAACGGCCAGGCGATGCTACAGGTGATTTTGA 2477  
Db ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2478 AACCGTTTCGTTGGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAATCCGCTTT 2537  
Db AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759

2538 TTAATCAACCCTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2598 TGTCTTATCATCACAGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657  
Db CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGACGACTGGCAGAGGCGCAGACT 2717  
Db ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCGAG 2777  
Db GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2778 AAGATCAGACGGGAGCTTAAACAGGCTTACTGTGGATTCTTCTCGTCTGGAGTTCTTCA 2837  
Db LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGTGATGCT 2897  
Db GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2898 AGACTTAAAGCCTTAATACAGATCCTGCTGGCAGCTGTGTAGTACGAGCAGACGTGGTTAT 2957  
Db ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017  
Db PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 3018 GAGAGGAACTGACTGTTGGAGAGTATATATAAGCTGCTGTACGATATTACAATGAAGAA 3077  
Db GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 3078 TGCAGAACTGCTCCACCCCGGACCAACATCAAGCTTTATCCATTCATATACCATGCA 3137  
Db CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 3138 GTTGAGTCTGTACACAGACCACCAACCCGCGGACAAAGGACGGG 3185  
Db ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975  
RESULT 7  
AAU76012  
ID AAU76012 standard; protein; 976 AA.  
XX  
AC AAU76012;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX  
OS Homo sapiens.  
XX  
PN US6333148-B1.  
XX  
PD 25-DEC-2001.  
XX  
PF 30-APR-1999; 99US-00302812.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;



XX WPI; 2002-153820/20.  
DR N-PSDB; ABK14494.  
XX  
PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
PS Claim 3; Col 55-60; 80pp; English.  
XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage,  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the human PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX  
SQ Sequence 976 AA;  
  
Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4635.50 Matches: 874  
Percent Similarity: 92.83% Conservatives: 32  
Best Local Similarity: 89.55% Mismatches: 69  
Query Match: 64.01% Indels: 1  
DB: 5 Gaps: 1  
  
US-09-302-812-1 (1-4070) x AAU76012 (1-976)  
QY 258 ATGAGTGGCGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCGCTGCA 317  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
QY 318 ACTTCTCCGGCGCGCCCTCGGACGCGCGGAGCTTCCCGGCGAGGCGGCGCTCTC 377  
Db 21 ThrSer--ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 378 GATTCCAAAGGACGCTCCGGTGCAGTTCAGGTCCCGCGCTCTCGTCAGGTCGCGCCCTG 437  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 438 GGCGGCGGCGGACAGCACCAAGGAATCAAGACAGATTGAATCAGAAAGTTGCATAGTAA 497  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGATTGAATCAGAAAGTTGCATAGTAA 557  
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 558 GAAACAAACATACAGAGAGAATCCATGATGAGTTCTGTACAAAAGATAACTTTTAT 617  
Db 100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
QY 618 CAACATAACATGAAAATAGAAAATGTTTCTCAGCTAGGTTTGTGATAAGTCACCAAGTT 677  
Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGGCGCTATGTGTAACTGGCAGAA 737  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 798 GAGCAGTTCAGTAATGCTAATGCTCATGTCAGTCGTCGCCCAAGGATGATCATCAGTGACACA 857

Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 858 AATAGTGAGGAGAGTAGAGATATATCAGCAGTTTGTGACACATGTAAAGCTTGCGAATGCA 917  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 918 AAGCAGACGATGGAAGATGAACAGGCGCAGAGAGCCAGAGCCACCAAGTGTGGCAAG 977  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 978 GCTTGCCATCTCTGAGAGCCCTGTGACGGGTGTGACGAGGAGGAGACAGCTGGTGTCC 1037  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
QY 1038 GAGAGCCCTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 1098 AGATTGATAGACAAAGAAAGTAGTCTAGGAAATCTCCTCCATTGAGAAAGAAAGTGAA 1157  
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299  
QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCAAAAAATAGTTGTGAGGATTCAGAGCAGAT 1217  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 1218 GAAGAGACAAGTCCAGGTTTGTGATGAACAGAGATAGCAGTTCTGCTCAACAGCAAAAT 1277  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
QY 1278 AAACCTTCAAGGTTCCAAACAGAGAGAGCTGACTGAGTTGAGGAAGCGGCTCTCTGCT 1337  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1338 AAGGGAGGTGAGATTGATTACATTTCATTTCAATTTGAAGGAGGAGAGATCGAGCTGGAATG 1397  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1398 AATGATGTGAATGCCAAACACCTGGAAGTACTTCTAGCTGAATGTAGAGTGCAGAAAT 1457  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTTCTAAATCACAGATCATTTTCATGAGAGTGCCCAA 1517  
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1518 GCAGAGCAACAAAGAAAGAACAAATGTGAATGAAACATCAAAAGAACAGAAAGGAGATC 1577  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1578 CCTAAATACATTCCACCTCACCTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGGGATCCGGCTGCCCTCCCTTGAGACCATCTGCCAAT 1697  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1698 CACACAGTCACTATTCGGGTAGATCTTTTTCGAATAGGAGAGATTCTTAAACCTTTCCCA 1757  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1758 ACACATTTTAAAGATTGTGGGACAAACAAGCATGTTAAGATGCCCTGTTCAGAACAAAC 1817  
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1818 TTGTACCTCTGGAAGATGAGAATGTTGAGCGAGCTGCAGGAGCGCGGTGGGAACACTCAT 1877  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1878 CAGACTGCATTCTCAACAGGCTCACTCGCCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937



|          |          |  |      |
|----------|----------|--|------|
| Db       | 540      | GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys   | 559  |
| QY       | 1938     | TACAATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTGATTGATTTCTGGGATAAG  | 1997 |
| Db       | 560      | TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys   | 579  |
| QY       | 1998     | GTACTAGAAGACGAGAAGCTCAACACTTGATCAGTCCATCTTGCTGCTGATATGGTGAAA   | 2057 |
| Db       | 580      | ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys   | 599  |
| QY       | 2058     | ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCCAATACCCTCTGAAACAGAGATG   | 2117 |
| Db       | 600      | IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet   | 619  |
| QY       | 2118     | AATCATTTCCATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC   | 2177 |
| Db       | 620      | AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe   | 639  |
| QY       | 2178     | TGCACGTTTCCACGCGCAATGCCAAGATGAATCAGAGTATTCAGTTATCCAGATATT  | 2237 |
| Db       | 640      | CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle   | 659  |
| QY       | 2238     | AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAGGAACACAGAGAAGCTTAAACGCTC  | 2297 |
| Db       | 660      | AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu   | 679  |
| QY       | 2298     | TTCTGCTACTTTAGAGAGTCACAGAGAAACCCACTGGTGGTGATTCACAAGA   | 2357 |
| Db       | 680      | PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg   | 699  |
| QY       | 2358     | CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT   | 2417 |
| Db       | 700      | GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis   | 719  |
| QY       | 2418     | GTCACTTACGAAGGTACCATAGAGGAACCGCCAGGCGCATGCTACAGGTGGATTTTGCA  | 2477 |
| Db       | 720      | ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla   | 739  |
| QY       | 2478     | AACCGTTTCGTTGGAGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAGAAATCGGCTTT  | 2537 |
| Db       | 740      | AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe  | 759  |
| QY       | 2538     | TTAATCAACCTGAGTTGATTTGTTTACGGCTCTTCACTGAGGTGCTGGATCACAATGAA  | 2597 |
| Db       | 760      | LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu   | 779  |
| QY       | 2598     | TGCTTATCATCACAGGTACTGAGCAGTACAGTGAATACAGGCTATGCCGGAACATAC  | 2657 |
| Db       | 780      | CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr   | 799  |
| QY       | 2658     | CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACAGCTGGCAGAGGGCGCACT   | 2717 |
| Db       | 800      | ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr   | 819  |
| QY       | 2718     | GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG   | 2777 |
| Db       | 820      | GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu   | 839  |
| QY       | 2778     | AAGATCAGCGGAGCTTAACAAGGCTTACTGTGGATTTCTCGTCTCGAGTTTCTTCA   | 2837 |
| Db       | 840      | LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer   | 859  |
| QY       | 2838     | GAGAACCTGTCTGCAGTGGCTACAGGAACCTGGGCTGTGGTCTTTGGGGGTGATGCT  | 2897 |
| Db       | 860      | GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla   | 879  |
| QY       | 2898     | AGACTAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGACGAGACGTGGTTTAT   | 2957 |
| Db       | 880      | ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr  | 899  |
| QY       | 2958     | TTTCCCTTTGGGACTCAGAACTGATGAGACATTTACAGCATGATCATCTCTCACT  | 3017 |
| Db       | 900      | PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr   | 919  |
| QY       | 3018     | GAGAGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGTACGATATTACAATGAAGAA   | 3077 |
| Db       | 920      | GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu   | 939  |
| QY       | 3078     | TGCAGAAACTGCTCCACCCCGGACGACATCAAGCTTTATCCATTATATACCATGCA   | 3137 |
| Db       | 940      | CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla   | 959  |
| QY       | 3138     | GTTGAGTCTGTACACAGACCACCAACCAGCCGGGACAAAGACGGGG   | 3185 |
| Db       | 960      | ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly   | 975  |
| RESULT 8 |          |  |      |
| ID       | ABG72279 | standard; protein; 976 AA.   |      |
| XX       | AC       | ABG72279;  |      |
| XX       | DT       | 13-MAR-2003 (first entry)  |      |
| XX       | DE       | Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.   |      |
| XX       | KW       | Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression; cellular response; DNA damage; neoplastic disorder inducing agent; oxidative stress; neoplastic disorder; myocardial infarction;  |      |
| KW       | KW       | vascular stroke; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; Huntington's disease; inborn genetic error; reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytosolic; neuroprotective; neurotropic; antiparkinsonian; cardiac; vasotrophic; anticonvulsant; cerebroprotective; enzyme.  |      |
| XX       | OS       | Homo sapiens.  |      |
| XX       | PN       | US2002132328-A1.   |      |
| XX       | PD       | 19-SEP-2002.   |      |
| XX       | PF       | 09-OCT-2001; 2001US-00973451.  |      |
| XX       | PR       | 01-MAY-1998; 98US-0083768P.  |      |
| PR       | PR       | 30-APR-1999; 99US-00302812.  |      |
| XX       | PA       | (JACO/) JACOBSON M K.  |      |
| PA       | PA       | (JACO/) JACOBSON E L.  |      |
| PA       | PA       | (AMEJ/) AME J.   |      |
| XX       | PI       | (LINW/) LIN W.   |      |
| XX       | PI       | Jacobson MK, Jacobson EL, Ame J, Lin W;  |      |
| XX       | DR       | WPI; 2003-155895/15.   |      |
| DR       | DR       | N-PSDB; ABX14478.  |      |
| XX       | PT       | New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose) glycohydrolase activity, for preventing, treating, or ameliorating a disease condition, e.g. neoplastic disorder, myocardial infarction or vascular stroke.   |      |
| PT       | PT       | Claim 28; Fig 16; 86pp; English.   |      |
| XX       | PS       | The present invention relates to the isolation of poly(ADP-ribose) glycohydrolase (PARG) from several species, and the polynucleotide sequences encoding them. Methods for inhibiting PARG expression or overexpressing PARG are also disclosed. PARG is involved in the cellular response to DNA damage, and is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. The polynucleotide sequences encoding PARG and PARG modulators are useful for preventing, treating, or ameliorating diseases such as neoplastic disorders, myocardial infarction, vascular stroke, neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's disease), inborn genetic errors, reperfusion following ischaemia, aging, and neurotoxicity. The polynucleotide sequences are also useful in gene therapy. The methods are useful for identifying an agent that modulates |      |

CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents human PARG enzyme

XX  
SQ Sequence 976 AA;

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4635.50 Matches: 874  
Percent Similarity: 92.83% Conservative: 32  
Best Local Similarity: 89.55% Mismatches: 69  
Query Match: 64.01% Indels: 1  
DB: 6 Gaps: 1

US-09-302-812-1 (1-4070) x ABG72279 (1-976)

QY 258 ATGAGTGGGGCCCGGTGTGAGCCCTGCACCAACGACCCCGCTGGGACGCGCTGCA 317  
|||:|||||  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
|||:|||||  
QY 318 ACTTCTCCGGCCCGCTCGGACGCCCGGAGCTTCCCGGCAGGCAGAGGCGCTCCTC 377  
|||:|||||  
Db 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39  
|||:|||||  
QY 378 GATTCGAAGGACGCTCCGGTGCAGTTTCAGGGTCCCGCGTCTCTCGTCAGGCTGGCCCTG 437  
|||:|||||  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
|||:|||||  
QY 438 GGCCGGCGGACAGCACCGGAGGACGCCACCTCTCTTGTCTTCAACAGAGACTATA 497  
|||:|||||  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
|||:|||||  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGATAGTAAA 557  
|||:|||||  
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
|||:|||||  
QY 558 GAAACAACAATACAGAGAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTTAT 617  
|||:|||||  
Db 100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
|||:|||||  
QY 618 CAACATAACATGGAATAATTAGAAAATGTTTCTCAGCTAGGTTTGTATAAGTCACCAATT 677  
|||:|||||  
Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
|||:|||||  
QY 678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAT 737  
|||:|||||  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
|||:|||||  
QY 738 GAAGGSCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
|||:|||||  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
|||:|||||  
QY 798 GAGCAGTTCAGTAATGTCTGATCAGTCAGTCTGTCGCCCAAGAGGATGATCAGATGACACA 857  
|||:|||||  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
|||:|||||  
QY 858 AATAGTCAGGAGAGTAGAGATAATCAGCAGTTTGTGACACATGTAAGCTTGCGAATGCA 917  
:|||:|||||  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
|||:|||||  
QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAGCCAGAGAGTGTGGCAAG 977  
|||:|||||  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
|||:|||||  
QY 978 GCTTGCCATCTGCAGAACGCTGTGCGAGGTGTGAGGAGGAGGAGCAGACGCTGGTGTC 1037  
:|||:|||||  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro 259  
|||:|||||  
QY 1038 GAGAGCCCTTGTGGACACTGGCTCTGAGGATGTGGTACTGGACTGAAATAATGCCAAC 1097  
:|||:|||||  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
|||:|||||  
QY 1098 AGATTGAATAGACAGAAAGTAGTCTAGGAAATCTCTCCATTGTAGAAAGAAAGTGAA 1157  
|||:|||||

Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
|||:|||||  
QY 1158 CCTGAGTCACCAATGGATGTAGATAATCCAAAAATAGTTGTCAAGGATTCAGAAAGCAGAT 1217  
|||:|||||  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
|||:|||||  
QY 1218 GAAGAGACAAAGTCCAGGTTTGTATGAACAGGAAGATAGCAGTTCTGCTCAAAACAGCAAT 1277  
|||:|||||  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
|||:|||||  
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAAAGCTGACACTGAGTTGAGGAAGCGGCTCTCTGCT 1337  
|||:|||||  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
|||:|||||  
QY 1338 AAGGAGGTGAGATTTCGATTACATTTCCAAATTTGAAGGAGGAGAGATCGAGTGGAAATG 1397  
|||:|||||  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
|||:|||||  
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
|||:|||||  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
|||:|||||  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTTCTAAATCACAGATCATTTTCATGAGAGTGCCTCAAA 1517  
|||:|||||  
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
|||:|||||  
QY 1518 GCAGAGGACAAAAGAAAAGAAACAATGTGAATGAAATGAAACATCAAGAACAGAAAGAGATC 1577  
|||:|||||  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
|||:|||||  
QY 1578 CCTAAATACATTCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTCTATTGAG 1637  
|||:|||||  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
|||:|||||  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGATCCGGCTGCCCTTCCCTTGAGACCATCTGCCAAT 1697  
|||:|||||  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
|||:|||||  
QY 1698 CACACAGTGACTATTTCGGGTAGATCTTTTGGCAATAGAGAAAGTTCTTAAACCTTCCCA 1757  
|||:|||||  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
|||:|||||  
QY 1758 ACACATTTTAAAGATTTCGGGACAAACAAGCATGTTAAGATGCCTTGTTCAGAAACAAAC 1817  
|||:|||||  
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
|||:|||||  
QY 1818 TTGTACCTGTGGAAGATGAGAAATGGTGAGCGAGCTGACGCGCGGTGGGAACCTCAAT 1877  
|||:|||||  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
|||:|||||  
QY 1878 CAGACTGCACCTTCTCAACAGGTCACCTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
|||:|||||  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
|||:|||||  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTGATTTCTGGGATAAG 1997  
|||:|||||  
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
|||:|||||  
QY 1998 GTACTAGAAGAGCAGAGCTCAACACTTGTATCAGTCCACTTTCCTGATATGGTGAAA 2057  
|||:|||||  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
|||:|||||  
QY 2058 ATTGCACTCTGTCTGCCAATATTGTACCCAGCCCAATACCCTCCTGAAACAGAAAGATG 2117  
|||:|||||  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
|||:|||||  
QY 2118 AATCATTCCTCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTTAATGCTTTCTTC 2177  
|||:|||||  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
|||:|||||  
QY 2178 TGCACGTTTCCAGCAGCAATGCCAAGATGAAATCAGAGTATTCAGTATTCAGATATT 2237  
|||:|||||

Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659

QY 2238 AACTTCAATCGGTTGTTGAAGGACGTTTCATCAAGGAAACAGAGAAGCTTAAACGCTC 2297

Db 660 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 679

QY 2298 TTCTGCTACTTTAGAGAGTCACAGAGAAAAAACCACCTGGTGGTGACATTCAACAAGA 2357

Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699

QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCTGACTCGACTGCAT 2417

Db 700 GlnSerLeuGluAspPheProGluTyrGluArgCysGluLysProLeuThrArgLeuHis 719

QY 2418 GTCACCTTACGAAGGTACCATAGAGGAACCGCCAGGGCATGCTACAGGTGGATTGCA 2477

Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739

QY 2478 AACCGTTTCGTTGGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAATCCGCTTT 2537

Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759

QY 2538 TTAATCAACCTCAGTTGATTGTTTCAACGCTCTTCACTGAGGTGCTGGATCAACAATGAA 2597

Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

QY 2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657

Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799

QY 2658 CGCTGGGCGCGGAGCCATGAAGACAGGAGCGGAAAGGACGACTGGCAGAGGCGCAGACT 2717

Db 800 ArgTyrSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819

QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCGATTGTGCCCCGAG 2777

Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

QY 2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGATTTCTTCTGCTGGAGTTCTTCA 2837

Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

QY 2838 GAGAACCTGTCTGAGTGGTACAGGAAACTGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897

Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla 879

QY 2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957

Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899

QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTATACAGCATGCATACATTCCTCACT 3017

Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

QY 3018 GAGAGGAACTGACTGTTGGAGAAATATATAAGCTGCTGTACGATATTACAATGAAGAA 3077

Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939

QY 3078 TGCAGAACTGCTCCACCCCGGACAGACATCAAGCTTTATCCATTATATACCATGCA 3137

Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

QY 3138 GTTGAGTCTGTACACAGACCCACACCGCCGGGACAAAGACGGGG 3185

Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975

RESULT 9

AAE25631

ID AAE25631 standard; protein; 968 AA.

XX

AC

XX

XX

DT 04-NOV-2002 (first entry)

XX DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).

XX KM Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;

KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;

KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;

KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;

KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;

XX KM antisenese therapy.

OS Mus musculus.

XX US6395543-B1.

PN 28-MAY-2002.

XX 23-FEB-2000; 2000US-005111507.

XX 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

PA Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

DR N-PSDB; AAD42083.

DR New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX Claim 3; Col 63-68; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The antisense therapeutic are useful to treat neoplastic disorders and conditions caused by genotoxic oxidative stress e.g., cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. The invention is useful in gene therapy and antisense therapy.

XX The present sequence is murine PARG

XX Sequence 968 AA;

SQ

Alignment Scores:

Pred. No.: 0 Length: 968

Score: 4332.50 Matches: 820

Percent Similarity: 89.86% Conservative: 57

Best Local Similarity: 84.02% Mismatches: 90

Query Match: 59.82% Indels: 9

DB: 5 Gaps: 5

US-09-302-812-1 (1-4070) x AAE25631 (1-968)

QY 258 ATGAGTGGGGCCCGCTGTGAGCCCTGCACCAAGGACCCCGTGGACGCGCTGCA 317

Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLys---AlaArgTyrGlyAlaGly 19

QY 318 ACTTCTCCGCGCGCTCGGACGCGCGGAGCTTCCCGGAGGAGGCGGCTCCTC 377

Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39

QY 378 GATTCCAAGGACGCTCCGCTGTCAGTTTCAGGGTCCCGCGCTCCTCGTCAGGCTGCGCCTG 437

Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59



QY 438 GGCGGGGGGACAGCAGCAGCGAGCGCCACCTCTCTTGTGTTTCAAAACAGAAAGACTATA 497  
Dp 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACATAAAGGAATCAAGACAGTGAATCAGAAAAGTTTGCATAGTAAA 557  
Dp 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
QY 558 GAAACAAACAATACAGAGAAAGAAATCCATGATGAGTCTGTACAAAAGATAACTTTTAT 617  
Dp 97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116  
QY 618 CAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGGTTTGTGATAAGTCACCAGTT 677  
Dp 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
QY 678 GAAAAAGGTACACAGTATTGAAGCAGCATCAGACTCGCGCTATGTGTAAAGTGGCAGAAT 737  
Dp 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
QY 738 GAAGGGCCACACTCAGAACCGCTTTTGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Dp 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
QY 798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCAGAGTGACACA 857  
Dp 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
QY 858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTCTTTTGACACATGTAAAGCTTGGCAATGCA 917  
Dp 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
QY 918 AAGCAGCAGATGGAAGATGAACAGGGCAGAGAAAGCCAGAACCCAGCAGAGTGTGGCAAG 977  
Dp 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233  
QY 978 GCTTGCCATCCTGCAGAAAGCCTGTGAGGTTGTGAGGTTGTGAGGAGGAGACAGCAGTGTGTC 1037  
Dp 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253  
QY 1038 GAGAGCCCTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAATAATGCCAAC 1097  
Dp 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
QY 1098 AGATTGAATACAGAAAGTAGTCTAGGAATCTCTCCATTTTGAAAGAAAGTGAA 1157  
Dp 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProPheGluLysGluSerGlu 293  
QY 1158 CCTGAGTCACCAATGGATGTAGATAATCCAAAATAGTTGTGTCAGGATTGAGAGCAGAT 1217  
Dp 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
QY 1218 GAAGAGACAAGTCCAGGTTTGTATGAACAGGAAGATAGCAGTTCTGCTCAACAGCAAAT 1277  
Dp 314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332  
QY 1278 AAACCTTCAAGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAAGCGGTCTCTGCT 1337  
Dp 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
QY 1338 AAGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397  
Dp 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371  
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCTGAATGTAGAGTGCAAAAT 1457  
Dp 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerSerLeuAsnValGluCysArgSer 391  
QY 1458 TCTAAGCAACATGGGAGAAAGGATCTCTAAATCAGATCATTTTCATGAGAGTGGCCAAA 1517  
Dp 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411  
QY 1518 GCAGAGGACAAAAGAAAGAACAAATGTGAAATGAAACATCAAGAAACAGAAAGGAAGATC 1577

Db 412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
QY 1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAAGAAATGGCTTGGAATCCTATTGAG 1637  
Dp 432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCAATCTGCCAAT 1697  
Dp 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
QY 1698 CACACAGTGACTATTCCGGGTAGATCTTTTCCGAATAGGAGAAGTTCCATAACCTTTCCCA 1757  
Dp 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491  
QY 1758 ACACATTTTAAAGATTGTGGGACAAACAAGCATGTAAAGATGCCTTGTTCAGAAACAAAAC 1817  
Dp 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
QY 1818 TTGTACCCTGTGGAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAATCATT 1877  
Dp 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531  
QY 1878 CAGACTGCACCTTCTCAACAGGCTCAACACTTGTATCAGTCCATCTGCTGATATGTTGTAAG 1937  
Dp 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTGATTTGTTGGGATAAG 1997  
Dp 552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571  
QY 1998 GTACTAGAAAGAACGAGCAAGCTCAACACTTGTATCAGTCCATCTGCTGATATGTTGTAAG 2057  
Dp 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
QY 2058 ATTGCACTCTGTCTGCCAAATATTGTACCCAGCCAAATACCACTCTCTGAAACAGAAAGATG 2117  
Dp 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611  
QY 2118 AATCATTTCCATCACAAATGTCAAGGAACAGATTGCCAGTCTTTTAGTAATGCTTTCTTC 2177  
Dp 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
QY 2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGATTATCCAGATATT 2237  
Dp 632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651  
QY 2238 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACAGAGAAACAGAGAACTTAAACGCTC 2297  
Dp 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671  
QY 2298 TTCTGCTACTTTAGAAAGATCACAGAGAGAAACCCACTGGGTGGTGACATTCACAAGA 2357  
Dp 672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691  
QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417  
Dp 692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711  
QY 2418 GTCACCTTACGAAGTACCATAGAAAGAAACCGCCAGGGCATGCTACAGGTGGATTTTGCA 2477  
Dp 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
QY 2478 AACCGTTTCGTTGGAGGTGGTGTAAACCAGTGCAGGACTTGTGCAAGAAAGAAATCCGCTTT 2537  
Dp 732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751  
QY 2538 TTAATCAACCCCTGAGTTGATTTTTCACGGCTTTCCTGAGGTGCTGGATCACAATGAA 2597  
Dp 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771  
QY 2598 TGTCTTATCATCACAGGTAAGTACAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657



Db 772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791  
QY 2658 CGCTGGGCCCGGAGCATGAAGACAGAGCGAAGAGGACGACTGGCAGAGCGCAGACT 2717  
Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811  
QY 2718 GAGATCGTCGCCATCGAGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCGAG 2777  
Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831  
QY 2778 AAGATCAGACGGGAGCTTAAACAGGCTTACTGTGGATTCTTCTCGTCTGGAGTTCTTCA 2837  
Db 832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
QY 2838 GAGAACCTGTCTGAGTGGCTACAGAACTGGGGCTGGTGGCTTGGGGGTGATGCT 2897  
Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871  
QY 2898 AGACTAAAGCCCTTAATACAGATCCTGGCAGCTGTGTAGCTGAGCGAGACGTGGTTTAT 2957  
Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 891  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCTCTCACT 3017  
Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
QY 3018 GAGAGGAACTGACTGTTGGAGAGTATATAAGCTGTCTGCTACATATTACATGAAGAA 3077  
Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931  
QY 3078 TGCAGAACTCTCCACCCCGGACCGACATCAAGCTTTATCCATTCATATACCATGCA 3137  
Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
QY 3138 GTTGAATCTGTACACAGACCACCAACCGCGGACAAAGGACGGGG 3185  
Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967

RESULT 10  
AAU76022  
ID AAU76022 standard; protein; 968 AA.  
XX  
AC AAU76022;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
OS Mus musculus.  
XX  
PN US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-163240/21.  
DR N-PSDB; ABK14933.  
XX

Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
treating neoplastic and neurological disorders, heart attack and stroke.  
Claim 2; Col 63-70; 81pp; English.  
The present invention relates to a new poly(ADP-ribose) glycohydrolase  
(PARG) protein which catalyses release of ADP-ribose from an ADP  
(adenosine diphosphate)-ribose polymer. The PARG molecule of the  
invention is useful for generating antibodies and can be inhibited or  
activated for diagnosing and treating neoplastic disorders such as  
adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
and related conditions. PARG levels may be enhanced to suppress DNA  
repair and increase the cell's susceptibility to chemotherapy drugs.  
Antagonists of PARG are administered to treat or prevent neoplastic  
disorders. The present amino acid sequence represents the mouse PARG  
protein of the invention. This protein is one of several PARG proteins  
(AAU76020-AAU76024) of the invention  
SQ Sequence 968 AA;  
Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 59.82% Indels: 9  
DB: 5 Gaps: 5  
US-09-302-812-1 (1-4070) x AAU76022 (1-968)  
QY 258 ATGAGTGGCGGCGGCTGTGAGCCCTGCACCAAGCAGCCCGCTGGAGCGCGCTGCA 317  
Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLys---AlaArgTrpGlyAlaAlaGly 19  
QY 318 ACTTCTCCGCGCGGCTCGGACCGCGCGGAGCTTCCCGGCGAGGAGCGCGCTCCTC 377  
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
QY 378 GATTCCAGGACGCTCCGTTGCGAGTTCAGGGTCCCGCGCTCCTCGTCAGGCTGCGCCCTG 437  
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
QY 438 GGCGGCGGCGGACAGCAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 497  
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGACACTAAAGGAATCAAGACAGTGAATCAGAAAGTTGCATAGTAA 557  
Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
QY 558 GAAACAAACAATACAAAGAGAGATCCATGATGAGTCTGTACAAAAGATAACTTTTAT 617  
Db 97 GluAsnAsnThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyr 116  
QY 618 CAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGTGTGATAAGTCACGAGTT 677  
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
QY 678 GAAAAGGTACACAGTATTTGAAGCAGCATCAGACTCGGCTATGTGTAGTGCAGAAAT 737  
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
QY 738 GAAGGCGCACACTCAGAACGGCTTTTGGAAAGTGTAACCTCCAGCGGTAACTCTGTACCA 797  
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
QY 798 GAGCAGTTTCAAGTATGTAATGTCGATCAGTCGTCCTCCCAAGGATGATCAGTGACACA 857  
Db 177 LysGlnLeuSerAsnAlaAlaIleGlyGlnSerProHisThrAspAspHisSerAspThr 196



Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931  
QY 3078 TGCAGAACTGCTCCACCCCGGACGACATCAAGCTTTATCCATTATACCATGCA 3137  
Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
QY 3138 GTTGAGTCCTGTACACAGACCACCAACCGCGGGACAAAGGACGGG 3185  
Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967

RESULT 11

AAU76013

ID AAU76013 standard; protein; 968 AA.

XX AC AAU76013;

XX 08-MAY-2002 (first entry)

XX Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.

XX Mus musculus.

XX US6333148-B1.

XX 25-DEC-2001.

XX 30-APR-1999; 99US-00302812.

XX 01-MAY-1998; 98US-0093768P.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-153820/20.

XX N-PSDB; ABK14495.

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.

XX Claim 3; Col 63-68; 80pp; English.

XX The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the mouse PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention

XX Sequence 968 AA;

SQ Alignment Scores:

Pred. No.: 0 Length: 968  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 59.82% Indels: 9  
DB: 5 Gaps: 5

US-09-302-812-1 (1-4070) x AAU76013 (1-968)  
QY 258 ATGAGTCCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317  
Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLys---AlaArgTrpGlyAlaAlaGly 19  
QY 318 ACTTCTCCGGCCGGCCCTCGGACGCCCGGAGCTTCCCGCGCAGGCAGAGCGCGTCTC 377  
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
QY 378 GATTCCAAAGGACGCTCCGGTGTGAGTTCAGGGTCCCGCGCTCCTCGTCAGGCTGCGCCCTG 437  
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
QY 438 GGCCGGCGGGACAGCACCCGAGGACGCCCCACCTCTCTTGTTCATAAAGACTATA 497  
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTGAATCAGAAAGTTTGCATAGTAA 557  
Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
QY 558 GAAACAACAATAACAAGAGAAGATCCATGATGAGTCTCTGTACAAAAGATACTTTTAT 617  
Db 97 GluAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116  
QY 618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTGTGATAAGTCACCAGTT 677  
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
QY 678 GAAAAAGGTACACAGTATTGTAAGCAGCATCAGACTCGGGCTATGTGTAAGTGGCAGAAT 737  
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
QY 738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
QY 798 GAGCAGTTCAAGTAATGCTAATGTCGATCAGTCTCGTCCCAAGAGATGATCAGTGCACACA 857  
Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
QY 858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTCTTTTGACACATGTAAGCTTGCAGATGCA 917  
Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
QY 918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAAGCCAGAACCCAGAAAGTGTGGCAAG 977  
Db 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233  
QY 978 GCTTGCCATCCTGCAGAACCCCTGTGCGGGTGTTCAGCAGGAGGAGACAGACGTGGTGCC 1037  
Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253  
QY 1038 GAGAGCCCTTGTCTGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAATGCCAAC 1097  
Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
QY 1098 AGATTGAATAGACAAAGAAAGTAGTCTAGGAAATTTCTCTCCATTTGAGAAAGAAAGTGAA 1157  
Db 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProPheGluLysGluSerGlu 293  
QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTGAGGATTCAGAAAGCAGAT 1217  
Db 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
QY 1218 GAAGAGACAAAGTCCAGGTTTGTGATGAACAGGAAGATAGCAGTTCTGCTCAAAACAGCAAAT 1277  
Db 314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332  
QY 1278 AAACCTTCAAGGTTCCCAACCAAGAGAGCTGACACTGAGTTGAGGAAAGCGGCTCTCTGCT 1337



Db 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
QY 1338 AAGGGAGGTGAGATTGATTACATTTCCAAATTTGAAGGAGGAGAGACTGCGTGAATG 1397  
Db 353 LysGlySerGluValArgLeuHisPheGlnPheGlu--GlyGluAsnAsnAlaGlyThr 371  
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
Db 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerLeuAsnValGluCysArgSer 391  
QY 1458 TCTAAGCAACATGGGAGAAAAGGATTCTAAAATCACAGATCATTTTCATGAGAGTGCCCAAA 1517  
Db 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411  
QY 1518 GCAGAGGACAAAAGAAAAGAACAAATGTGAATGAAATCAAAATCAAAAGAACAGAAAGGAGATC 1577  
Db 412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
QY 1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637  
Db 432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGGATCCGGCTGCCCTCCCTTGAGACCATCTGCCAAT 1697  
Db 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
QY 1698 CACACAGTGACTATTCCGGGTAGATCTTTTTCGGAATAGGAGAAGTTCTCTAAACCTTTCCCA 1757  
Db 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491  
QY 1758 ACACATTTTAAAGATTTGTGGGACAAACAGCATGTTAAAGATGCCTTGTTCAGAACAAAAAC 1817  
Db 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
QY 1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGGAGCTGCAGGACGCCGGTGGGAACCTATT 1877  
Db 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531  
QY 1878 CAGACTGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937  
Db 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997  
Db 552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571  
QY 1998 GTACTAGAAGAGCAGAGCTCAACACTTGTATCAGTCCACTTGCCTGATATGGTGAAG 2057  
Db 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
QY 2058 ATTGCACCTCTGTCTGCCAAATATTGTACCCAGCCAAATACCACTCCTGAAACAGAGATG 2117  
Db 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611  
QY 2118 AATCATTCATCACAAATGTCACGAAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177  
Db 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
QY 2178 TGCAAGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATCCAGTTATCCAGATATT 2237  
Db 632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTrpProAspIle 651  
QY 2238 AACTTCAATCGGTTGTTGAAGGACGTTTCATCAAGGAAACAGAGAAAGCTTAAACGCTC 2297  
Db 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671  
QY 2298 TTCTGCTACTTTAGAAGAGTCCAGAGAAAAAACCCACTGGGTGGTGACATTCAACAAGA 2357  
Db 672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691  
QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAAATCCTCGACTCGACTGCAT 2417  
Db 692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

QY 2418 GTCACCTTACGAAGGTACCATAGAGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477  
Db 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
QY 2478 AACCGTTTCGTTGGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAATCCGCTTT 2537  
Db 732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluIleArgPhe 751  
QY 2538 TTAATCAACCTGAGTTGATTGTTTTCACGCTCTTCACTGAGGTGCTGATCACAATGAA 2597  
Db 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771  
QY 2598 TGTCTTATCATCAGAGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAAACATAC 2657  
Db 772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791  
QY 2658 CGCTGGGCCCCGAGCCATGAAGACAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2717  
Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCACTTGTGCCCCGAG 2777  
Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831  
QY 2778 AAGATCAGACGGGAGCTTAAACAGGCTTAACTGTGGATTCTTCGTCCTCGAGTTTCTTCA 2837  
Db 832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
QY 2838 GAGAACCTGTCTGAGTGGCTACAGGAAACCTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897  
Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871  
QY 2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGTAGCTGAGCGAGACCTGGTTTAT 2957  
Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 891  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017  
Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
QY 3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAGCTGCTACGATGCTACGATATTAACATGAAGA 3077  
Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931  
QY 3078 TGCAGAAACTGCTCCACCCCGGACCAACATCAAGCTTTATCCATTCATATACCATGCA 3137  
Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
QY 3138 GTTGAGTCTGTACACAGACCAACCAACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3185  
Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967

RESULT 12  
ABG72280  
ID ABG72280 standard; protein; 968 AA.  
XX AC ABG72280;  
XX DT  
XX 13-MAR-2003 (first entry)  
DE Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
KW Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
KW cytosolic; neuroprotective; neurotropic; antiparkinsonian; cardiant;  
KW vasotrophic; anticonvulsant; cerebroprotective; enzyme.  
OS Mus musculus.



XX US2002132328-A1.  
PN 19-SEP-2002.  
XX 09-OCT-2001; 2001US-00973451.  
PF 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2003-155895/15.  
XX N-PSDB; ABX14479.  
DR  
DR  
XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX Claim 28; Fig 16; 86pp; English.  
PS  
XX The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents murine PARG enzyme  
XX  
SQ Sequence 968 AA;  
Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 59.82% Indels: 9  
DB: 6 Gaps: 5  
US-09-302-812-1 (1-4070) x ABG72280 (1-968)  
QY 258 ATGAGTGGGCGGCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGAGCGCGCTGCA 317  
Db 1 MetSerAlaGlyProGlyTyrProGlyThrLys---AlaArgTyrGlyAlaAlaGly 19  
QY 318 ACTTCTCCGCGGCGGCTCGGACCGCGGAGCTTCCCGCGGCGGAGCGCGCTCCTC 377  
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
QY 378 GATTCCAGGACGCTCCGGTGCAGTTCAGGGTCCCGCGCTCCTCGTCAGGCTGCGCCCTG 437  
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
QY 438 GGCCGGGCGGACAGCAGCCAGGCGGCGCCACCTCTCTTTTCAACAGAGACTATA 497  
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrile 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTGTGAATCAGAAAGTTTGCATAGTAA 557

Db 80 ThrThrTyrMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
QY 558 GAAACCAACAATACAGAGAGAATCCATGATGAGTTCTGTACAAAAGATAACTTTAT 617  
Db 97 GluAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116  
QY 618 CAACATAACATGGAAAAATAGAAAAATGTTTCTCAGCTAGGTTTGTATAAGTCCAGTT 677  
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
QY 678 GAAAAAGGTACACAGTATTGAAAGCAGCATCAGACTCGGCTATGTGTAAAGTGGCAGAT 737  
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTyrGlnAsn 156  
QY 738 GAAGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGTATCCA 797  
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
QY 798 GAGCAGTTTCAATGCTAATGTCGATCAGTCGTCCCAAGGATGATCAGAGTGCACACA 857  
Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspHisSerAspThr 196  
QY 858 AATAGTGAGGAGTAGAGATAATCAGCAGTGTGACACATGTAAAGCTTCCGAATGCA 917  
Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProLysLeuAlaAsnThr 216  
QY 918 AAGCAGCAGTGAAGATGAACAGGCGCAGAGAGCCAGAGCCACAGAGTGTGGCAAG 977  
Db 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233  
QY 978 GCTTGCATCTCAGAAAGCTGTGAGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1037  
Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253  
QY 1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097  
Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
QY 1098 AGATTGAATAGACAAAGAAAGTAGTCTAGAAATTTCTCTCCATTGAGAAAGAAAGTAA 1157  
Db 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProPheGluGlyGluSerGlu 293  
QY 1158 CCTGAGTCAACCAATGATGTAGATAATTCACAAATAATGTTGTGAGGATTCAGAGCAGAT 1217  
Db 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
QY 1218 GAAGAGACAAGTCCAGGTTTGTGTAACAGAGAGATAGCAGTTCTGCTCAAAACAGCAAT 1277  
Db 314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332  
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAGGAGGAGGAGGAG 1337  
Db 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
QY 1338 AAGGGAGGTGAGATTCGATTACATTTCCATTTGAAGGAGGAGAGAGTCCGAGTGGAAATG 1397  
Db 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371  
QY 1398 AATGATGTGAATGCCAAACAGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
Db 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTTCTAAATCAGATCATTTTCATGAGAGTCCCAAA 1517  
Db 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411  
QY 1518 GCAGAGGACAAAGAAAGAAACAATGTGAATGAACATCAAAAGAACAGAAAGAGATC 1577  
Db 412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
QY 1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637

Db 432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGGATCCGGTCCCTCCCTTGAGACCATCTGCCAAT 1697  
Db 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
QY 1698 CACACAGTGACTATTTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCCTAACCTTTCCCA 1757  
Db 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491  
QY 1758 ACACATTTTAAAGATTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAAACAAAAC 1817  
Db 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
QY 1818 TTGTACCCTGTGGAGATGAGAATGGTGAGCGGAGCTGCAGGCAGCCGGTGGGAACCTATT 1877  
Db 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531  
QY 1878 CAGACTGCACCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGATGCTATTCTGAAG 1937  
Db 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551  
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTTGATTGATTCTTCGGATAAG 1997  
Db 552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571  
QY 1998 GTACTAGAAGAAGCAGAGCTCAACACTTGTATCAGTCCACTTTCCTGATATGGTGAAA 2057  
Db 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
QY 2058 ATTGCACCTCTGTCTGCCAAATATTGTACCCAGCCATACCACCTCTCTGAAACAGAAGATG 2117  
Db 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611  
QY 2118 AATCATTTCCATCACAATGTCAAGAAATGTTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177  
Db 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
QY 2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGTTATCCAGATATT 2237  
Db 632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651  
QY 2238 AACTTCAATCGTGTGTTGAGAGAGCTTCATCAAGGAACAGAGAAAGCTTAAACGCTC 2297  
Db 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671  
QY 2298 TTCTGCTACTTTAGAGAGTCAAGAGAAAAAACCCTGCTGGTGGTGACATTCACAAGA 2357  
Db 672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691  
QY 2358 CAGAGCTTTGAAGATTTCACAGAGTGGGAAAGATGTGAAAACCTCTGACTCGACTGCAT 2417  
Db 692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711  
QY 2418 GTCACTACGAAGGTACCATAGAGAAACGGCCAGGGCAGTCTACAGGTGGATTTCGA 2477  
Db 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
QY 2478 AACCGTTTCGTTGGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537  
Db 732 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751  
QY 2538 TTAATCAACCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
Db 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771  
QY 2598 TGTCTTATCATCAGGTACTGAGCAGTACAGTGAATACACAGGCTATCCGAAACATAC 2657  
Db 772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791  
QY 2658 CGCTGGGCCCGAGCCATGAAGACAGGAGCGAAAGGACGACTGGCAGAGCGCAGACT 2717  
Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgCysThr 811

QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCACTTTGTGCCCCGAG 2777  
Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831  
QY 2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTCTTCGTCCTCGAGTTTCTTCA 2837  
Db 832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
QY 2838 GAGAACCTGTCTGCAGTGGCTACAGGAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897  
Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871  
QY 2898 AGACTAAAGCCTTAATACAGATCCTGGCAGCTGTGTAGCTGAGCGAGACGCTGTTTAT 2957  
Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 891  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTACAGCATGCATACATTCCTCACT 3017  
Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
QY 3018 GAGAGGAAACTGACTGTGTGGAGAAGTATATAGCTGTCTACGATATACATGAAGAA 3077  
Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuArgTyrTyrAsnGluGlu 931  
QY 3078 TGCAGAAACTGCTCCACCCCGGACCAACCAACCGCGGACAAAGGACGCGG 3137  
Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
QY 3138 GTTGAGTCCGTACACAGACCAACCAACCGCGGACAAAGGACGCGG 3185  
Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967  
RESULT 13  
ABBS9491  
ID ABBS9491 standard; protein; 768 AA.  
XX AC ABBS9491;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 5265.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX PI WPI; 2001-656860/75.  
XX DR N-PSDB; ABL03594.  
XX DR  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell  
XX PT interactions.  
XX PS Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.  
XX PS  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and





DT 04-NOV-2002 (first entry)  
XX Fruit fly poly adenosine diphosphate-ribose glycohydrolase (PARG).  
DE  
XX  
KW Poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
XX ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytotstatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy; fruit fly.  
XX  
OS Drosophila melanogaster.  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-535641/57.  
DR N-PSDB; AAD42084.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Example 13; Col 71-76; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX  
SQ The present sequence is fruit fly PARG  
SQ Sequence 768 AA;  
  
Alignment Scores:  
Pred. No.: 9.7e-80 Length: 768  
Score: 1063.50 Matches: 244  
Percent Similarity: 57.31% Conservative: 93  
Best Local Similarity: 41.50% Mismatches: 198  
Query Match: 14.69% Indels: 53  
DB: 5 Gaps: 15  
  
US-09-302-812-1 (1-4070) x AAE25632 (1-768)  
  
QY 1506 AGAGTGCCCAAGCAGAGGACAAAGAAAGAAACAATGTGAATGAACAATCAAGAACA 1565  
Db 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGlu----- 62  
  
QY 1566 GAAAGGAAGATCCCTAAATACATTCACCTCACCTT-----TCTCCAGATAAGAAATGG 1619  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspAspSerTirp 74  
  
QY 1620 CTTGGAACTCTATTGAGGAGATG-----AGGAGATGCCAAGGTGTGGATCCGGCTG 1673  
Db 75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

QY 1674 CCTCCCTTGAGACCATCTGCCAATCACACAGTGACTATTTCGGGTAGATCTTTTGGGAATA 1733  
Db 95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGln-----LeuProIle 111  
  
QY 1734 GGAGAAAGTTCTTAAACCTTTCCCAACACATTTTAAAGATTGTGGCAACAAGCATGTT 1793  
Db 112 ArgGluThrPro---ProArgProTyrLysSerProGlyLysTyrAspSerGluHisVal 130  
  
QY 1794 AAGATGCCTTGTTCAGAACAAACTTGTACCTGTGGAAGATGAGATGGTGAGCGAGCT 1853  
Db 131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150  
  
QY 1854 GCAGGCAGCGGTGGGAATCATTCACTGACTGCTTCTCAACAGGCTCACTGGCCCCAG 1913  
Db 151 IleAspPheArgTrpGluMetIleGluArgAlaLeuGlnProIleLysThrCysGlu 170  
  
QY 1914 AACCTGAGGATGCTATTCTGAAGTACAATGTGGCATAATTCTAAGAAATGGGACTTTACA 1973  
Db 171 GluLeuGlnAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190  
  
QY 1974 GCTTTGATTGATTCTGGGATAAGGTACTAGAAGACAGAGCTCAACACTTGTATCAG 2033  
Db 191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210  
  
QY 2034 TCCATCTTGCTGATATGGTGAAATTTGCACCTCTGTCTGCCAAATATTTGTACCCAGCCA 2093  
Db 211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230  
  
QY 2094 ATACCACCTCTGAAACAGAGATGAATCATCTCCATCACAATGTACAGGACAGATTGCC 2153  
Db 231 ValProLeuLeuHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSer 250  
  
QY 2154 AGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGACGCAATGCC---AAGATGAAA 2210  
Db 251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270  
  
QY 2211 TCAGAGTATTCAGTTATCCAGATATTAACCTCAATCGGTTGTTTGAAGGACGTTTCATCA 2270  
Db 271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290  
  
QY 2271 AGGAAACCAGAGAGCTTAAACGCTCTTCTGCTACTTTAGAGAGTCT-----ACAGAG 2324  
Db 291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310  
  
QY 2325 AAAAAA-----CCCACTGGGTGGTGACATTCACAAGACAGAGT-----CTT 2366  
Db 311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330  
  
QY 2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCTG-----ACTCGACTGCATGTC 2420  
Db 331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350  
  
QY 2421 ACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTGCAAAAC 2480  
Db 351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370  
  
QY 2481 CGTTTCGTTGGAGGTGGTGAACCAAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTA 2540  
Db 371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390  
  
QY 2541 ATCAACCCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGATGT 2600  
Db 391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410  
  
QY 2601 CTTATCATCACAGGTACTGACAGTACAGTGAATACACAGGCTATCCGAAACATACCCG 2660  
Db 411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430  
  
QY 2661 TGGGCGCGGAGCCATGAAGACAGGAGCGAAAGGACGACGTCGCAGAGGCGGACGACTGAG 2720  
Db 431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450





|    |      |  |      |  |  |      |
|----|------|--|------|--|--|------|
| Db | 231  | ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer   | 250  | 582                                      | LysLysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAsp | 598  |
| QY | 2154 | AGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGACGCAATGCC---AAGATGAAA   | 2210 | 3165                                     | CAGCCGGGACAAAGGACGGGGGCC                                 | 3188 |
| Db | 251  | CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgAsnThrLeuLysArgLys      | 270  | 599                                      | ValProGlyGluGlyAlaSerAla                                 | 606  |
| QY | 2211 | TCAGAGTATCCAGTTATCCAGATATTAACCTCAATCGGTGTTGTTGAAGGACGTTTCATCA  | 2270 | Search completed: May 26, 2004, 16:45:16 |  |      |
| Db | 271  | SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro   | 290  | Job time : 297.648 secs                  |  |      |
| QY | 2271 | AGGAAACCCAGAGAGCTTAAACCGCTCTTCTGCTACTTTAGAAGAGTC-----ACAGAG    | 2324 |  |  |      |
| Db | 291  | AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu   | 310  |  |  |      |
| QY | 2325 | AAAAAA-----CCCACCTGGGTGGTGACATTCACAAGACAGAGT-----CTT           | 2366 |  |  |      |
| Db | 311  | ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro   | 330  |  |  |      |
| QY | 2367 | GAAGATTTCCAGAGTGGGAAAGATGTGAAAAAATCCTG-----ACTCGACTGCATGTC     | 2420 |  |  |      |
| Db | 331  | GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal   | 350  |  |  |      |
| QY | 2421 | ACTTACGAGGTACCATAGAAGGAAACGCCAGGGCATGCTACAGGTGGATTTTGCAAAAC    | 2480 |  |  |      |
| Db | 351  | AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn   | 370  |  |  |      |
| QY | 2481 | CGTTTCGTTGGAGGTGGTGTAAACCAGTGACGAGCTTGTGCAAGAGAAATCCGCTTTTA    | 2540 |  |  |      |
| Db | 371  | LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheVal      | 390  |  |  |      |
| QY | 2541 | ATCAACCCCTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT | 2600 |  |  |      |
| Db | 391  | IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgPropheGluAla   | 410  |  |  |      |
| QY | 2601 | CTTATCATCACAGTACTGACAGTACAGTGAATACACAGGCTATGCCGAAACATACCCGC    | 2660 |  |  |      |
| Db | 411  | LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu   | 430  |  |  |      |
| QY | 2661 | TGGGCCCGGAGCCATGAAGACAGGAGCGGAAAGGACGACTGGCAGAGGGCGCAGACTGAG   | 2720 |  |  |      |
| Db | 431  | TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla   | 450  |  |  |      |
| QY | 2721 | ATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAGAAG   | 2780 |  |  |      |
| Db | 451  | IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu   | 470  |  |  |      |
| QY | 2781 | ATCAGACGGGAGCTTAAAGGCTTACTGTGGATTCTTCGT-----                   | 2822 |  |  |      |
| Db | 471  | MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro   | 490  |  |  |      |
| QY | 2823 | CCTGGAGTTTCTTCAGAGAACCTGTCTGCTGAGTGGCTACAGGAAACTGGGGCTGTGGTGCC | 2882 |  |  |      |
| Db | 491  | ProGly-----ValAlaThrGlyAsnTrpGlyCysGlyAla                      | 502  |  |  |      |
| QY | 2883 | TTTGGGGTGATGCTAGACTAAAGCCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAG    | 2942 |  |  |      |
| Db | 503  | PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly   | 522  |  |  |      |
| QY | 2943 | CGAGACGTGTTTATTTTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATG    | 3002 |  |  |      |
| Db | 523  | ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMet      | 542  |  |  |      |
| QY | 3003 | CATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAGTATATAAGCTGCTGTACGA     | 3062 |  |  |      |
| Db | 543  | TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg   | 561  |  |  |      |
| QY | 3063 | TATTACAATGAAGAATGCAGAAACTGCTCCACCCCGGACCA-----                 | 3104 |  |  |      |
| Db | 562  | SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer   | 581  |  |  |      |
| QY | 3105 | GACATCAGCTTTATCCATTTCATATACCATGTCAGTTTGAGTCTCTGTACACAGACCACAC  | 3164 |  |  |      |

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 0.352479 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description          |
|------------|-------|-------------|--------|-------|----------------------|
| 1          | 53    | 100.0       | 11     | 4     | US-09-302-812-13     |
| 2          | 53    | 100.0       | 11     | 4     | US-09-511-477-13     |
| 3          | 53    | 100.0       | 11     | 4     | US-09-511-507-13     |
| 4          | 53    | 100.0       | 968    | 4     | US-09-302-812-6      |
| 5          | 53    | 100.0       | 968    | 4     | US-09-511-477-6      |
| 6          | 53    | 100.0       | 968    | 4     | US-09-511-507-6      |
| 7          | 53    | 100.0       | 977    | 4     | US-09-302-812-2      |
| 8          | 53    | 100.0       | 977    | 4     | US-09-511-477-2      |
| 9          | 53    | 100.0       | 977    | 4     | US-09-511-507-2      |
| 10         | 52    | 98.1        | 976    | 4     | US-09-302-812-4      |
| 11         | 52    | 98.1        | 976    | 4     | US-09-511-477-4      |
| 12         | 52    | 98.1        | 976    | 4     | US-09-511-507-4      |
| 13         | 41    | 77.4        | 252    | 4     | US-09-198-452A-694   |
| 14         | 37    | 69.8        | 1323   | 1     | US-08-026-138E-4     |
| 15         | 37    | 69.8        | 1336   | 2     | US-08-231-193A-58    |
| 16         | 37    | 69.8        | 1336   | 2     | US-08-486-273A-58    |
| 17         | 37    | 69.8        | 1336   | 3     | US-08-940-086A-58    |
| 18         | 37    | 69.8        | 1336   | 4     | US-08-940-035A-58    |
| 19         | 37    | 69.8        | 1336   | 4     | US-08-935-105A-58    |
| 20         | 37    | 69.8        | 1336   | 4     | US-09-648-797-58     |
| 21         | 37    | 69.8        | 1336   | 4     | US-09-386-123-58     |
| 22         | 36    | 67.9        | 123    | 4     | US-09-328-352-5925   |
| 23         | 35    | 66.0        | 358    | 4     | US-08-858-207A-398   |
| 24         | 34    | 64.2        | 134    | 4     | US-09-489-039A-11592 |
| 25         | 34    | 64.2        | 409    | 4     | US-09-540-236-2952   |
| 26         | 34    | 64.2        | 1065   | 3     | US-09-412-545-2      |
| 27         | 33    | 62.3        | 88     | 4     | US-09-328-352-7192   |

|    |    |      |     |   |                      |                   |
|----|----|------|-----|---|----------------------|-------------------|
| 28 | 33 | 62.3 | 176 | 4 | US-09-107-532A-6622  | Sequence 6622, Ap |
| 29 | 33 | 62.3 | 313 | 4 | US-09-551-826D-14    | Sequence 14, Appl |
| 30 | 33 | 62.3 | 400 | 4 | US-09-543-681A-4587  | Sequence 4587, Ap |
| 31 | 33 | 62.3 | 452 | 4 | US-09-198-452A-853   | Sequence 853, App |
| 32 | 33 | 62.3 | 883 | 4 | US-09-489-039A-11249 | Sequence 11249, A |
| 33 | 32 | 60.4 | 169 | 2 | US-08-895-939-4      | Sequence 4, Appli |
| 34 | 32 | 60.4 | 169 | 3 | US-09-188-820-4      | Sequence 4, Appli |
| 35 | 32 | 60.4 | 172 | 4 | US-09-543-681A-7238  | Sequence 7238, Ap |
| 36 | 32 | 60.4 | 176 | 4 | US-09-543-681A-6922  | Sequence 6922, Ap |
| 37 | 32 | 60.4 | 247 | 4 | US-09-134-000C-6345  | Sequence 6345, Ap |
| 38 | 32 | 60.4 | 264 | 1 | US-08-188-582-26     | Sequence 26, Appl |
| 39 | 32 | 60.4 | 264 | 1 | US-08-646-715-26     | Sequence 26, Appl |
| 40 | 32 | 60.4 | 325 | 2 | US-09-018-576-3      | Sequence 3, Appli |
| 41 | 32 | 60.4 | 325 | 2 | US-09-018-576-12     | Sequence 12, Appl |
| 42 | 32 | 60.4 | 325 | 3 | US-09-248-137-3      | Sequence 3, Appli |
| 43 | 32 | 60.4 | 325 | 3 | US-09-248-137-12     | Sequence 12, Appl |
| 44 | 32 | 60.4 | 422 | 4 | US-09-252-991A-27513 | Sequence 27513, A |
| 45 | 32 | 60.4 | 501 | 4 | US-09-687-360-2      | Sequence 2, Appli |

ALIGNMENTS

RESULT 1  
US-09-302-812-13  
; Sequence 13, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 13  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-13

Query Match 100.0%; Score 53; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.00016;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||||  
Db 1 FLINPELIVSR 11

RESULT 2  
US-09-511-477-13  
; Sequence 13, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812

; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 13  
 ; LENGTH: 11  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 US-09-511-477-13

Query Match 100.0%; Score 53; DB 4; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.00016;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
 Db 1 FLINPELIVSR 11

RESULT 3

US-09-511-507-13  
 ; Sequence 13, Application US/095111507  
 ; Patent No. 6395543

GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,507  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 13  
 ; LENGTH: 11  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 US-09-511-507-13

Query Match 100.0%; Score 53; DB 4; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.00016;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
 Db 1 FLINPELIVSR 11

RESULT 4

US-09-302-812-6  
 ; Sequence 6, Application US/09302812B  
 ; Patent No. 6333148

GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/302,812B  
 ; CURRENT FILING DATE: 1999-04-30  
 ; EARLIER APPLICATION NUMBER: 60/083,768  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6  
 ; LENGTH: 968  
 ; TYPE: PRT

; ORGANISM: Mus musculus  
 ; FEATURE:  
 US-09-302-812-6

Query Match 100.0%; Score 53; DB 4; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 0.036;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
 Db 751 FLINPELIVSR 761

RESULT 5

US-09-511-477-6  
 ; Sequence 6, Application US/09511477  
 ; Patent No. 6337202

GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 US-09-511-477-6

Query Match 100.0%; Score 53; DB 4; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 0.036;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
 Db 751 FLINPELIVSR 761

RESULT 6

US-09-511-507-6  
 ; Sequence 6, Application US/09511507  
 ; Patent No. 6395543

GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,507  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 US-09-511-507-6

Query Match 100.0%; Score 53; DB 4; Length 968;



Best Local Similarity 100.0%; Pred. No. 0.036; Indels 0; Gaps 0;  
Matches 11; Conservative 0; Mismatches 0;

QY 1 FLINPELIVSR 11  
| | | | | | | | | |  
Db 751 FLINPELIVSR 761

RESULT 7

US-09-302-812-2  
; Sequence 2, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-2

Query Match 100.0%; Score 53; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
| | | | | | | | | |  
Db 760 FLINPELIVSR 770

RESULT 8

US-09-511-477-2  
; Sequence 2, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-2

Query Match 100.0%; Score 53; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
| | | | | | | | | |

Db 760 FLINPELIVSR 770

RESULT 9

US-09-511-507-2  
; Sequence 2, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-2

Query Match 100.0%; Score 53; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
| | | | | | | | | |  
Db 760 FLINPELIVSR 770

RESULT 10

US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.058;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
| | | | | | | | | |  
Db 759 FLINPELIVSR 769

RESULT 11  
US-09-511-477-4

; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-477-4

Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.058;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||||:|  
Db 759 FLINPELIISR 769

# RESULT 12

US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-507-4

Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.058;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||||:|  
Db 759 FLINPELIISR 769

# RESULT 13

US-09-198-452A-694  
; Sequence 694, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Grifffais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 694  
; LENGTH: 252  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-694

Query Match 77.4%; Score 41; DB 4; Length 252;  
Best Local Similarity 80.0%; Pred. No. 1.9;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIVS 10  
|||||:|  
Db 48 FLINPERVVS 57

# RESULT 14

US-08-026-138E-4  
; Sequence 4, Application US/08026138E  
; Patent No. 5502166  
; GENERAL INFORMATION:  
; APPLICANT: Masayoshi MISHINA  
; TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nishiohata Residence 1-107  
; STREET: 5214, Nishiohata-machi  
; CITY: Niigata-shi  
; STATE: Niigata-ken  
; COUNTRY: JAPAN  
; ZIP: 951  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS v.5  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/026,138E  
; FILING DATE: 26-FEB-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 39563/1992  
; FILING DATE: 26-FEB-1992  
; APPLICATION NUMBER: JP 173155/1992  
; FILING DATE: 30-JUN-1992  
; APPLICATION NUMBER: JP 215017/1992  
; FILING DATE: 12-AUG-1992  
; APPLICATION NUMBER: JP 303878/1992  
; FILING DATE: 13-NOV-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hamburg, C.Bruce  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: F-4551  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 986-2340  
; TELEFAX: (212) 953-7733  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1323 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single strand  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: mouse  
; TISSUE TYPE: brain  
; PUBLICATION INFORMATION:  
; AUTHORS: Masayoshi MISHINA  
; TITLE: NOVEL PROTEINS AND GENES CODING THE SAME

*Don't take out  
being published*

/ RELEVANT RESIDUES IN SEQ ID NO: 4: FROM 1 to 1323  
US-08-026-138E-4

Query Match 59.8%; Score 37; DB 1; Length 1323;  
Best Local Similarity 66.7%; Pred. No. 88;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLINPELIV 9  
||:|:|:|  
Db 378 FLVNPSLVV 386

RESULT 15  
US-08-231-193A-58  
; Sequence 58, Application US/08231193A  
; Patent No. 5849895  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,193A  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,459  
; FILING DATE: 20-APR-1993  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1336 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-231-193A-58

Query Match 69.8%; Score 37; DB 2; Length 1336;  
Best Local Similarity 66.7%; Pred. No. 89;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLINPELIV 9  
||:|:|:|  
Db 381 FLVNPSLVV 389

Search completed: May 26, 2004, 18:49:32  
Job time : 1.35248 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 0.917775 Seconds  
(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 1149313 seqs, 278921704 residues 1149313  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description          |
|------------|-------|---------------|--------|-------|----------------------|
| 1          | 53    | 100.0         | 11     | 9     | US-09-973-451-13     |
| 2          | 53    | 100.0         | 968    | 9     | US-09-973-451-6      |
| 3          | 53    | 100.0         | 977    | 9     | US-09-973-451-2      |
| 4          | 52    | 98.1          | 976    | 9     | US-09-973-451-4      |
| 5          | 42    | 79.2          | 119    | 12    | US-10-424-599-168770 |
| 6          | 42    | 79.2          | 546    | 12    | US-10-425-114-60000  |
| 7          | 41    | 77.4          | 249    | 12    | US-10-282-122A-54709 |
| 8          | 41    | 77.4          | 252    | 15    | US-10-289-762-694    |
| 9          | 38    | 71.7          | 59     | 12    | US-10-424-599-215648 |
| 10         | 37    | 69.8          | 1336   | 9     | US-09-945-901-58     |
| 11         | 37    | 69.8          | 1336   | 13    | US-10-007-747-58     |
| 12         | 37    | 69.8          | 1336   | 14    | US-10-038-937-58     |
| 13         | 36    | 67.9          | 107    | 12    | US-10-424-599-183741 |
| 14         | 35    | 66.0          | 449    | 12    | US-10-282-122A-73824 |
| 15         | 35    | 66.0          | 527    | 15    | US-10-369-493-1735   |

|    |    |      |      |    |                      |                   |
|----|----|------|------|----|----------------------|-------------------|
| 16 | 35 | 66.0 | 621  | 14 | US-10-156-761-10025  | Sequence 10025, A |
| 17 | 35 | 66.0 | 820  | 14 | US-10-174-677-30     | Sequence 30, Appl |
| 18 | 35 | 66.0 | 828  | 14 | US-10-174-677-90     | Sequence 90, Appl |
| 19 | 35 | 66.0 | 932  | 14 | US-10-174-677-89     | Sequence 89, Appl |
| 20 | 35 | 66.0 | 932  | 14 | US-10-174-677-91     | Sequence 91, Appl |
| 21 | 34 | 64.2 | 21   | 16 | US-10-663-896-17     | Sequence 17, Appl |
| 22 | 34 | 64.2 | 39   | 12 | US-10-424-599-231144 | Sequence 231144,  |
| 23 | 34 | 64.2 | 72   | 12 | US-10-424-599-176489 | Sequence 176489,  |
| 24 | 34 | 64.2 | 86   | 12 | US-10-424-599-183835 | Sequence 183835,  |
| 25 | 34 | 64.2 | 98   | 12 | US-10-424-599-183834 | Sequence 183834,  |
| 26 | 34 | 64.2 | 100  | 12 | US-10-424-599-162739 | Sequence 162739,  |
| 27 | 34 | 64.2 | 138  | 10 | US-09-896-580A-8     | Sequence 8, Appli |
| 28 | 34 | 64.2 | 169  | 10 | US-09-896-580A-3     | Sequence 3, Appli |
| 29 | 34 | 64.2 | 211  | 12 | US-10-282-122A-47127 | Sequence 47127, A |
| 30 | 34 | 64.2 | 275  | 12 | US-10-424-599-169986 | Sequence 169986,  |
| 31 | 34 | 64.2 | 307  | 12 | US-10-425-114-58003  | Sequence 58003, A |
| 32 | 34 | 64.2 | 407  | 12 | US-10-282-122A-63173 | Sequence 63173, A |
| 33 | 34 | 64.2 | 408  | 15 | US-10-369-493-3561   | Sequence 3561, Ap |
| 34 | 34 | 64.2 | 449  | 15 | US-10-369-493-20063  | Sequence 20063, A |
| 35 | 34 | 64.2 | 451  | 12 | US-10-282-122A-72152 | Sequence 72152, A |
| 36 | 34 | 64.2 | 651  | 15 | US-10-369-493-21499  | Sequence 21499, A |
| 37 | 34 | 64.2 | 761  | 12 | US-10-114-270-108    | Sequence 108, App |
| 38 | 34 | 64.2 | 1065 | 9  | US-09-771-161A-239   | Sequence 239, App |
| 39 | 33 | 62.3 | 58   | 12 | US-10-424-599-211138 | Sequence 211138,  |
| 40 | 33 | 62.3 | 59   | 9  | US-09-864-761-38589  | Sequence 38589, A |
| 41 | 33 | 62.3 | 241  | 12 | US-10-424-599-255332 | Sequence 255332,  |
| 42 | 33 | 62.3 | 269  | 12 | US-10-282-122A-71472 | Sequence 71472, A |
| 43 | 33 | 62.3 | 295  | 12 | US-10-282-122A-53386 | Sequence 53386, A |
| 44 | 33 | 62.3 | 428  | 12 | US-10-282-122A-54992 | Sequence 54992, A |
| 45 | 33 | 62.3 | 428  | 15 | US-10-312-273-93     | Sequence 93, Appl |

ALIGNMENTS

RESULT 1

US-09-973-451-13  
; Sequence 13, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA2 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 13  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-13

Query Match 100.0%; Score 53; DB 9; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.0025;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
| | | | |  
Db 1 FLINPELIVSR 11

RESULT 2



US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

Query Match 100.0%; Score 53; DB 9; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||  
Db 751 FLINPELIVSR 761

RESULT 3  
US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

Query Match 100.0%; Score 53; DB 9; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||  
Db 760 FLINPELIVSR 770

RESULT 4  
US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4

Query Match 98.1%; Score 52; DB 9; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.54;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||  
Db 759 FLINPELIVSR 769

RESULT 5  
US-10-424-599-168770  
; Sequence 168770, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 168770  
; LENGTH: 119  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(119)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_123414C.1.pap  
US-10-424-599-168770

Query Match 79.2%; Score 42; DB 12; Length 119;  
Best Local Similarity 63.6%; Pred. No. 4;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||  
Db 71 FLVQPELVVNR 81



US-10-424-599-215648

Query Match 71.7%; Score 38; DB 12; Length 59;  
Best Local Similarity 45.5%; Pred. No. 11;  
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
||:||||:  
Db 22 FLEPEMVITR 32

## RESULT 10

US-09-945-901-58  
; Sequence 58, Application US/09945901  
; Patent No. US20020161215A1

## GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.  
Ellis, Steven B.  
Liaw, Chen W.  
Lu, Chin-Chun

TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR

SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR

NUMBER OF SEQUENCES: 63

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA

COUNTRY: U.S.A.

ZIP: 92037

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/945,901

FILING DATE: 24-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/940,035

FILING DATE: <Unknown>

APPLICATION NUMBER: US 08/052,449

FILING DATE: 20-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 6362-9383E

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999

TELEFAX: 619-238-0062

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 1336 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 58:

US-09-945-901-58

## Query Match

Best Local Similarity 69.8%; Score 37; DB 9; Length 1336;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 9

||:||||:  
Db 381 FLVNPFLV 389

## RESULT 11

US-10-007-747-58

; Sequence 58, Application US/10007747

; Publication No. US20020161193A1

## GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.  
Ellis, Steven B.  
Liaw, Chen W.  
Lu, Chin-Chun

TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR

SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR

NUMBER OF SEQUENCES: 63

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe

STREET: 4250 Executive Square, 7th Floor

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,747

FILING DATE: 07-Dec-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/648,797

FILING DATE: 28-Aug-2000

APPLICATION NUMBER: US/08/940,086A

FILING DATE: 29-SEPT-97

APPLICATION NUMBER: US 08/231,193

FILING DATE: 20-APR-1994

APPLICATION NUMBER: US 08/052,449

FILING DATE: 20-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24735-9383C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 450-8400

TELEFAX: (619) 450-8499

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 1336 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 58:

US-10-007-747-58

## Query Match

Best Local Similarity 69.8%; Score 37; DB 13; Length 1336;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 9

||:||||:  
Db 381 FLVNPFLV 389

## RESULT 12

US-10-038-937-58

; Sequence 58, Application US/10038937

; Publication No. US20030013866A1

## GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.

Lu, Chin-Chun

TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR

SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR

NUMBER OF SEQUENCES: 63

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe

STREET: 4250 Executive Square, 7th Floor

CITY: La Jolla

STATE: CA

COUNTRY: U.S.A.

ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/038,937  
FILING DATE: 18-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/935,105  
FILING DATE: 29-SEPT-97  
APPLICATION NUMBER: US 08/231,193  
FILING DATE: 20-APR-1994  
APPLICATION NUMBER: US 08/052,449  
FILING DATE: 20-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9383D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1336 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 58:  
US-10-038-937-58

Query Match 69.8%; Score 37; DB 14; Length 1336;  
Best Local Similarity 66.7%; Pred. No. 5.1e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 9  
Db 381 FLVNPPLV 389

RESULT 13  
US-10-424-599-183741  
; Sequence 183741, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 183741  
; LENGTH: 107  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(107)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_136932C.1.pep  
US-10-424-599-183741

Query Match 67.9%; Score 36; DB 12; Length 107;  
Best Local Similarity 60.0%; Pred. No. 48;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIVS 10  
Db 94 FMINPTLIIN 103

RESULT 14  
US-10-282-122A-73824  
; Sequence 73824, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73824  
; LENGTH: 449  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-282-122A-73824

Query Match 66.0%; Score 35; DB 12; Length 449;  
Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 LINPELIVSR 11  
Db 119 IVNPELLAAR 128

RESULT 15  
US-10-369-493-1735  
; Sequence 1735, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.



APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 1735  
LENGTH: 527  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1735

Query Match 66.0%; Score 35; DB 15; Length 527;  
Best Local Similarity 63.6%; Pred. No. 4.3e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|:|:|:|:|  
Db 309 FMISPRLVIR 319

Search completed: May 26, 2004, 19:19:02  
Job time : 1.91778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 0.176239 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID        | Description        |
|------------|-------|-------------|--------|--------------|--------------------|
| 1          | 40    | 75.5        | 316    | 1 SRG4 CAEEL | P54126 caenorhabdi |
| 2          | 39    | 73.6        | 314    | 1 NOD1_RHILP | P23718 rhizobium 1 |
| 3          | 38    | 71.7        | 150    | 1 DEF_CLOTE  | Q895Q2 clostridium |
| 4          | 37    | 69.8        | 295    | 1 Y188_CLOPE | P26832 clostridium |
| 5          | 37    | 69.8        | 1323   | 1 NME4_MOUSE | Q03391 mus musculu |
| 6          | 37    | 69.8        | 1323   | 1 NME4_RAT   | Q62645 rattus norv |
| 7          | 37    | 69.8        | 1336   | 1 NME4_HUMAN | O15399 homo sapien |
| 8          | 36    | 67.9        | 185    | 1 DEF2_NITEU | Q82tc8 nitrosomona |
| 9          | 36    | 67.9        | 320    | 1 FLIG_BUCBP | Q89az9 buchnera ap |
| 10         | 36    | 67.9        | 714    | 1 ZW10_DROGR | O44219 drosophila  |
| 11         | 35    | 66.0        | 169    | 1 DEF1_VIBCH | Q9kvn3 vibrio chol |
| 12         | 35    | 66.0        | 175    | 1 DEF1_RICCN | Q92iz1 rickettsia  |
| 13         | 35    | 66.0        | 175    | 1 DEF_RICPR  | Q9zdv8 rickettsia  |
| 14         | 35    | 66.0        | 527    | 1 IF2G_YEAST | P32481 saccharomyc |
| 15         | 35    | 66.0        | 839    | 1 Y422_MYCPN | P75175 mycoplasma  |
| 16         | 35    | 66.0        | 932    | 1 CDG8_HUMAN | Q9y5g5 homo sapien |
| 17         | 35    | 66.0        | 932    | 1 CDG9_HUMAN | Q9y5g4 homo sapien |
| 18         | 34    | 64.2        | 112    | 1 YFIA_ECOLI | P11285 escherichia |
| 19         | 34    | 64.2        | 169    | 1 DEF_HAEIN  | P44786 haemophilus |
| 20         | 34    | 64.2        | 201    | 1 DEF2_PROMM | Q7v5f9 prochloroco |
| 21         | 34    | 64.2        | 211    | 1 KAD_BORBU  | O51378 borrelia bu |
| 22         | 34    | 64.2        | 215    | 1 TD01_SULME | Q55060 sulfolobus  |
| 23         | 34    | 64.2        | 245    | 1 YNFI_RHOCA | P17435 rhodobacter |
| 24         | 34    | 64.2        | 246    | 1 MCT1_MERUN | P50340 meriones un |
| 25         | 34    | 64.2        | 270    | 1 PPNK_LACLA | Q9c1j4 lactococcus |
| 26         | 34    | 64.2        | 285    | 1 Y011_MYCPN | P75098 mycoplasma  |
| 27         | 34    | 64.2        | 651    | 1 Y942_METJA | Q58352 methanococc |
| 28         | 34    | 64.2        | 726    | 1 YB1D_SCHPO | P87178 schizosacch |
| 29         | 34    | 64.2        | 1065   | 1 KDG1_HUMAN | Q75912 homo sapien |
| 30         | 33    | 62.3        | 130    | 1 NU5M_STREN | Q35832 strongyloce |
| 31         | 33    | 62.3        | 170    | 1 DEF1_VIBVU | Q8dde3 vibrio vuln |
| 32         | 33    | 62.3        | 170    | 1 DEF_PASMU  | P57948 pasteurella |
| 33         | 33    | 62.3        | 172    | 1 DEF1_VIBPA | Q87kds vibrio para |

|    |    |      |     |              |                    |
|----|----|------|-----|--------------|--------------------|
| 34 | 33 | 62.3 | 202 | 1 DEF2_VIBVY | Q7mgk6 vibrio vuln |
| 35 | 33 | 62.3 | 251 | 1 Y545_METJA | Q57965 methanococc |
| 36 | 33 | 62.3 | 305 | 1 EFTS_BRUME | Q8yhh5 brucella me |
| 37 | 33 | 62.3 | 313 | 1 MPR_EACSU  | P39790 bacillus su |
| 38 | 33 | 62.3 | 314 | 1 NOD2_RHITR | P32008 rhizobium t |
| 39 | 33 | 62.3 | 326 | 1 ACCD_SYNY3 | Q57417 synechocyst |
| 40 | 33 | 62.3 | 428 | 1 ENO_CHLPN  | Q9z7a6 chlamydia p |
| 41 | 33 | 62.3 | 479 | 1 ZW10_DROPS | O44218 drosophila  |
| 42 | 33 | 62.3 | 709 | 1 ETF2_FOWPV | Q9j562 fowlpox vir |
| 43 | 33 | 62.3 | 751 | 1 PSAA_CHLVU | P56341 chlorella v |
| 44 | 33 | 62.3 | 800 | 1 GUN_EACSI  | P06564 bacillus sp |
| 45 | 33 | 62.3 | 875 | 1 TRAC_ECOLI | P18004 escherichia |

ALIGNMENTS

RESULT 1  
SRG4 CAEEL  
ID SRG4 CAEEL STANDARD; PRT; 316 AA.  
AC P54126; Q95ZP2;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Serpentine receptor class gamma 4 (Srg-4 protein).  
GN SRG-4 OR T12A2.12.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Latreille P.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: Belongs to the C.elegans receptor-like protein srg family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U13019; AAK84568.2; -.  
PIR; T15557; T15557.  
DR WormPep; T12A2.12; CE33181.  
DR InterPro; IPR000609; Srg.  
DR PRINTS; PR00698; TMPTREINSRG.  
KW Transmembrane; Multigene family.  
FT TRANSMEM 21 41 POTENTIAL.  
FT TRANSMEM 52 72 POTENTIAL.  
FT TRANSMEM 88 108 POTENTIAL.  
FT TRANSMEM 140 160 POTENTIAL.  
FT TRANSMEM 188 208 POTENTIAL.  
FT TRANSMEM 229 249 POTENTIAL.  
FT TRANSMEM 259 279 POTENTIAL.  
SQ SEQUENCE 316 AA; 36716 MW; 05667CDFCC5E477D CRC64;

Query Match 75.5%; Score 40; DB 1; Length 316;  
Best Local Similarity 72.7%; Pred. No. 2.7;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
||| :|||  
Db 290 FLINPSPVSR 300

```

RESULT 2
NOD1_RHILP
ID NOD1_RHILP STANDARD; PRT; 314 AA.
AC P23718;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nodulation protein D I.
GN NODD1.
OS Rhizobium leguminosarum (biovar phaseoli).
OG Plasmid sym.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=385;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8002;
RX MEDLINE=91014692; PubMed=2215216;
RA Davis E.O., Johnston A.W.B.;
RT "Analysis of three nod genes in Rhizobium leguminosarum biovar
RT phaseoli; nodD1 is preceded by nOE, a gene whose product is secreted
RT from the cytoplasm."
RL Mol. Microbiol. 4:921-932(1990).
CC -!- FUNCTION: NODD REGULATES THE EXPRESSION OF THE NOD ABCFE GENES
CC WHICH ENCODE OTHER NODULATION PROTEINS. NODD IS ALSO A NEGATIVE
CC REGULATOR OF ITS OWN EXPRESSION. BINDS FLAVONOIDS AS INDUCERS.
CC -!- SIMILARITY: Contains 1 HTH LYSR-type DNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54214; CAA38126.1; -.
DR PIR; S11787; S11787.
DR InterPro; IPR000847; HTH_LysR.
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS50931; HTH_LYSR; 1.
KW Nodulation; Transcription regulation; DNA-binding; Activator;
KW Repressor; Plasmid.
FT DOMAIN 6 63 HTH_LYSR-TYPE.
FT DNA_BIND 23 42 H-T-H MOTIF (BY SIMILARITY).
FT SEQUENCE 314 AA; 35476 MW; A17DBC9965DE32B6 CRC64;
SQ SEQUENCE 314 AA; 35476 MW; A17DBC9965DE32B6 CRC64;

Query Match 73.6%; Score 39; DB 1; Length 314;
Best Local Similarity 72.7%; Pred. No. 4.3;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
DB 148 FLIFPELLMSR 158

RESULT 3
DEF_CLOTE
ID DEF_CLOTE STANDARD; PRT; 150 AA.
AC Q895Q2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).
GN DEF OR CTC01219.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]

Query Match 71.7%; Score 38; DB 1; Length 150;
Best Local Similarity 87.5%; Pred. No. 3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELI 8
DB 69 FLINPEII 76

RESULT 4
Y188_CLOPE
ID Y188_CLOPE STANDARD; PRT; 295 AA.
AC P26832;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein CPE0188.
GN CPE0188.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater."
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE OF 114-295 FROM N.A.
RC STRAIN=CPN50;
```





"Molecular characterization of the family of the N-methyl-D-aspartate receptor subunits.";  
J. Biol. Chem. 268:2836-2843(1993).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 2).  
TISSUE=Brain;  
RC MEDLINE=94206533; PubMed=7512349;  
RA Monyer H., Burnashev N., Laurie D.J., Sakmann B., Seeburg P.H.;  
"Developmental and regional expression in the rat brain and functional properties of four NMDA receptors.";  
Neuron 12:529-540(1994).  
[3]  
SEQUENCE FROM N.A. (ISOFORM 2).  
STRAIN=Sprague-Dawley; TISSUE=Brain;  
RA Boulter J., Pecht G.;  
Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NMDA receptor subtype of glutamate-gated ion channels with high calcium permeability and voltage-dependent sensitivity to magnesium. Mediated by glycine.  
CC -!- SUBUNIT: Heterodimer of an epsilon subunit and a zeta subunit.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=2;  
CC IsoId=Q62645-1; Sequence=Displayed;  
CC Name=1;  
CC IsoId=Q62645-2; Sequence=VSP\_000136;  
CC Note=No experimental confirmation available;  
CC -!- TISSUE SPECIFICITY: Expressed in brain, mainly in the subcortical region.  
CC -!- DEVELOPMENTAL STAGE: Already detected in embryonic stages, peaks at postnatal day 7, and decreases thereafter to adult levels.  
CC -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; D13213; BAA02500.1; -;  
DR EMBL; D13214; BAA02501.1; -;  
DR EMBL; L31611; AAC37646.1; -;  
DR EMBL; L31612; AAC37647.1; -;  
DR EMBL; U08260; AAA17833.1; -;  
DR PIR; I78557; I78557.  
DR HSSP; P19491; IGR2.  
DR InterPro; IPR001828; ANF receptor.  
DR InterPro; IPR001320; Ion\_glu\_receptor.  
DR InterPro; IPR001508; NMDA receptor.  
DR InterPro; IPR001311; SBP\_glu\_receptor.  
DR Pfam; PF01094; ANF\_receptor; 1.  
DR Pfam; PF00060; lig\_chan; 1.  
DR PRINTS; PR00177; NMDARECEPTOR.  
DR SMART; SM00079; PBPe; 1.  
KW Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium; Glycoprotein; Ionic channel; Magnesium; Alternative splicing.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 1323 GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 4.  
FT DOMAIN 28 580 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 581 601 1 (POTENTIAL).  
FT DOMAIN 602 623 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 624 644 2 (POTENTIAL).  
FT DOMAIN 645 653 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 654 674 3 (POTENTIAL).  
FT DOMAIN 675 841 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 842 862 4 (POTENTIAL).  
FT DOMAIN 863 1323 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 278 283 POLY-GLY.  
FT DOMAIN 905 913 POLY-PRO.

FT DOMAIN 1030 1035 POLY-ALA.  
FT DOMAIN 1197 1201 POLY-PRO.  
FT SITE 639 FUNCTIONAL DETERMINANT OF NMDA RECEPTORS (BY SIMILARITY).  
FT CARBOHYD 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 349 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 363 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 566 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 1265 CPTAAPTRLTGPSRHARCPHAAHWGPPPLPTASHRRHGG  
DLGTRRGSAHFSSLESEV -> RCPPPHRTGDTGAGTWAHA  
GALRISPAWSPRYDAAPATPTTAAAPSVSAGHGRGRAKWT  
GPSWVGKDRNGPGRTPPGAASCAPTFFALGEL (in isoform 1).  
FT /FTId=VSP\_000136.  
FT A -> V (IN REF. 3).  
FT P -> Q (IN REF. 3).  
FT G -> V (IN REF. 2; AAC37646).  
FT R -> P (IN REF. 2).  
FT R -> A (IN REF. 2).  
FT CONFLICT 305 305  
FT CONFLICT 635 635  
FT CONFLICT 974 974  
FT CONFLICT 1253 1253  
FT CONFLICT 1266 1267  
FT CONFLICT 1323 AA; 143100 MW; 40F7D60192579564 CRC64;  
SQ  
Query Match 69.8%; Score 37; DB 1; Length 1323;  
Best Local Similarity 66.7%; Pred. No. 50;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 FLINPELV 9  
Db 378 FLVNFSLVV 386  
-----  
RESULT 7  
NME4 HUMAN STANDARD; PRT; 1336 AA.  
ID NME4\_HUMAN  
AC O15399;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D) (EB11).  
DE GRIN2D.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal brain;  
RX MEDLINE=98149394; PubMed=9489750;  
RA Hess S.D., Daggett L.P., Deal C., Lu C.-C., Johnson E.C., RA Velicelebi G.;  
RT "Functional characterization of human N-methyl-D-aspartate subtype 1A/2D receptors.";  
RL J. Neurochem. 70:1269-1279(1998).  
CC -!- FUNCTION: NMDA receptor subtype of glutamate-gated ion channels with high calcium permeability and voltage-dependent sensitivity to magnesium. Mediated by glycine.  
CC -!- SUBUNIT: Heterodimer of an epsilon subunit and a zeta subunit.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC

EMBL; U77783; AAC15910.1; --  
HSSP; P19491; 1GR2.  
Genew; HGNC:4588; GRIN2D.  
MIM; 602717; --  
InterPro; IPR001828; ANF\_receptor.  
InterPro; IPR001320; Ion\_glu\_receptor.  
InterPro; IPR001508; NMDA\_receptor.  
InterPro; IPR001311; SBP\_glu\_receptor.  
Pfam; PF01094; ANF\_receptor; 1.  
Pfam; PF00060; lig\_chan; 1.  
PRINTS; PR00177; NMDARECEPTOR.  
SMART; SM00079; PBPe; 1.  
Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;  
KW Glycoprotein; Ionic channel; Magnesium.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 1336 GLUTAMATE [NMDA] RECEPTOR SUBUNIT  
FT EPSILON 4.  
FT DOMAIN 28 583 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 584 604 1 (POTENTIAL).  
FT DOMAIN 605 626 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 627 647 2 (POTENTIAL).  
FT DOMAIN 648 656 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 657 677 3 (POTENTIAL).  
FT DOMAIN 678 844 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 845 865 4 (POTENTIAL).  
FT DOMAIN 866 1336 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 281 286 POLY-GLY.  
FT DOMAIN 908 916 POLY-PRO.  
FT DOMAIN 1035 1040 POLY-ALA.  
FT DOMAIN 1209 1213 POLY-PRO.  
FT DOMAIN 1244 1247 POLY-ALA.  
FT SITE 642 FUNCTIONAL DETERMINANT OF NMDA  
RECEPTORS (BY SIMILARITY).  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 569 569 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1336 AA; 143558 MW; DECC545F3E416680 CRC64;

Query Match 69.8%; Score 37; DB 1; Length 1336;  
Best Local Similarity 66.7%; Pred. No. 51;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 9  
|||:|:|:  
DB 381 FLVNPSLV 389

RESULT 8  
DEF2\_NITEU STANDARD; PRT; 185 AA.  
AC Q82TC8;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Peptide deformylase 2 (EC 3.5.1.88) (PDF 2) (Polypeptide deformylase  
2).  
GN DEF2 OR NE1970.  
OS Nitrosomonas europaea.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;  
OC Nitrosomonadaceae; Nitrosomonas.  
OX NCBI\_TaxID=915;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19718 / IFO 14298;  
RX MEDLINE=22586410; PubMed=12700255;  
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,  
HAUSER L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,  
ARCIERO D.M., HOMMES N.G., WHITTAKER M.M., ARP D.J.;  
RT "Complete genome sequence of the ammonia-oxidizing bacterium and  
obligate chemolithoautotroph Nitrosomonas europaea.";  
RL J. Bacteriol. 185:2759-2773 (2003).

CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of  
CC newly synthesized proteins. Requires at least a dipeptide for an  
CC efficient rate of reaction. N-terminal L-methionine is a  
CC prerequisite for activity but the enzyme has broad specificity at  
CC other positions (By similarity).  
CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +  
CC methionyl peptide.  
CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).  
CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; BX321862; CAD85881.1; --  
CC HAMAP; MF\_00163; --; 1.  
CC InterPro; IPR000181; Fmet\_deformylase.  
CC Pfam; PF01327; Pep\_deformylase; 1.  
CC PRINTS; PR01576; PDEFORMYLASE.  
CC ProDom; PD003844; Pep\_deformylase; 1.  
CC TIGRFAMS; TIGR00079; pept\_deformyl; 1.  
KW Protein biosynthesis; Hydrolase; Iron; Complete proteome.  
FT ACT\_SITE 151 151 BY SIMILARITY.  
FT METAL 108 108 IRON (BY SIMILARITY).  
FT METAL 150 150 IRON (BY SIMILARITY).  
FT METAL 154 154 IRON (BY SIMILARITY).  
SQ SEQUENCE 185 AA; 20573 MW; C8E1A641426FDEC1 CRC64;

Query Match 67.9%; Score 36; DB 1; Length 185;  
Best Local Similarity 77.8%; Pred. No. 9.3;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LINPELIVS 10  
|||||:  
DB 90 LINPEIIAS 98

RESULT 9  
FLIG\_BUCBP STANDARD; PRT; 320 AA.  
AC Q89AZ9;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Flagellar motor switch protein Flig.  
GN FLIG OR BBP069.  
OS Buchnera aphidicola (subsp. Baizongia pistaciae).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Buchnera.  
OX NCBI\_TaxID=135842;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426901; PubMed=12522265;  
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,  
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,  
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;  
RT "Reductive genome evolution in Buchnera aphidicola.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586 (2003).  
CC -!- FUNCTION: FLIG IS ONE OF THREE PROTEINS (FLIG, FLIN, FLIM) THAT  
CC FORM A SWITCH COMPLEX THAT IS PROPOSED TO BE LOCATED AT THE BASE  
CC OF THE BASAL BODY. THIS COMPLEX INTERACTS WITH THE CHEY AND CHEZ  
CC CHEMOTAXIS PROTEINS, IN ADDITION TO CONTACTING COMPONENTS OF THE  
CC MOTOR THAT DETERMINE THE DIRECTION OF FLAGELLAR ROTATION (BY  
CC SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).  
CC -!- SIMILARITY: Belongs to the flig family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE014016; AAC26805.1; --  
DR InterPro; IPR000090; Flg\_Motor\_Flag.  
DR Pfam; PF01706; Flg-G; 1.  
DR PRINTS; PR00954; FLGMOTORFLIG.  
KW Chemotaxis; Flagellum; Flagellar rotation; Membrane;  
KW Complete proteome.  
SQ SEQUENCE 320 AA; 37070 MW; 82F74CDEE72DE287 CRC64;

Query Match 67.9%; Score 36; DB 1; Length 320;  
Best Local Similarity 88.9%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LINPELVS 10  
| | | | |  
DB 113 LINPEKVS 121

RESULT 10  
ZW10 DROGR STANDARD; PRT; 714 AA.  
AC O44219;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Centromere/kinetochore protein zw10 (Mitotic 15 protein).  
GN MIT(1)15 OR ZW10.

OS Drosophila grimshawi (Fruit fly) (Idiomysia grimshawi).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7222;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=97444363; PubMed=9298984;  
RA Starr D.A., Williams B.C., Li Z., Stenad-Moghadam B., Dawe R.K.,  
RA Goldberg M.L.;  
RT "Conservation of the centromere/kinetochore protein ZW10.";  
RL J. Cell Biol. 138:1289-1301(1997).  
CC -!- FUNCTION: Required for accurate chromosome segregation (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: EXCLUDED FROM THE NUCLEUS DURING INTERPHASE  
CC BUT MIGRATES INTO THE NUCLEAR ZONE DURING PROMETAPHASE. AT  
CC METAPHASE, FOUND IN A FILAMENTOUS STRUCTURE THAT MAY BE  
CC SPECIFICALLY ASSOCIATED WITH KINETOCHORE MICROTUBULES. AT  
CC ANAPHASE, FOUND AT OR NEAR KINETOCHORES OF SEPARATING CHROMOSOMES.  
CC AT THE BEGINNING OF TELOPHASE, BECOMES EXCLUDED AGAIN FROM THE  
CC NUCLEUS AND IS DISPERSED IN THE CYTOPLASM (BY SIMILARITY).  
CC -!- SIMILARITY: Belongs to the ZW10 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U54998; AAB88238.1; --  
DR FlyBase; FBgn0022892; Dgri\mit(1)15.  
KW Cell cycle; Meiosis; Mitosis; Nuclear protein; Centromere.  
SQ SEQUENCE 714 AA; 81262 MW; 62E14636E9D5B12D CRC64;

Query Match 67.9%; Score 36; DB 1; Length 714;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPEL 7  
| | | | |

Db 370 FLINPEL 376  
| | | | |  
RESULT 11  
DEF1 VIBCH STANDARD; PRT; 169 AA.  
AC Q9KVU3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Peptide deformylase 1 (EC 3.5.1.88) (PDF 1) (Polypeptide deformylase  
DE 1).  
GN DEF1 OR VC0046.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
RT cholerae.";  
RL Nature 406:477-483(2000).

CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of  
CC newly synthesized proteins. Requires at least a dipeptide for an  
CC efficient rate of reaction. N-terminal L-methionine is a  
CC prerequisite for activity but the enzyme has broad specificity at  
CC other positions (By similarity).  
CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +  
CC methionyl peptide.  
CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).  
CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE004095; AAF93224.1; --  
DR PIR; A82373; A82373.  
DR HSSP; P27251; 1DEF.  
DR TIGR; VC0046; --  
DR HAMAP; MF\_00163; --; 1.  
DR InterPro; IPR000181; Pep\_deformylase.  
DR Pfam; PF01327; Pep\_deformylase; 1.  
DR PRINTS; PR01576; PDEFORMYLASE.  
DR ProDom; PD003844; Pep\_deformylase; 1.  
DR TIGRFAMs; TIGR00079; pep\_deformyl; 1.  
KW Protein biosynthesis; Hydrolase; Iron; Complete proteome.  
FT METAL 91 91 IRON (BY SIMILARITY).  
FT METAL 133 133 IRON (BY SIMILARITY).  
FT ACT\_SITE 134 134 BY SIMILARITY.  
FT METAL 137 137 IRON (BY SIMILARITY).  
SQ SEQUENCE 169 AA; 19147 MW; 16DB00B08CA40FC7 CRC64;

Query Match 66.0%; Score 35; DB 1; Length 169;  
Best Local Similarity 70.0%; Pred. No. 13;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LINPELIVSR 11  
| | | | |

```
Db 73 LINPEIIEKR 82

RESULT 12
DEF1_RICCN STANDARD; PRT; 175 AA.
ID DEF1_RICCN Q92I21;
AC Q92I21;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Peptide deformylase 1 (EC 3.5.1.88) (PDF 1) (Polypeptide deformylase
1).
GN DEF1 OR RC0278.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Malish 7;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC newly synthesized proteins. Requires at least a dipeptide for an
CC efficient rate of reaction. N-terminal L-methionine is a
CC prerequisite for activity but the enzyme has broad specificity at
CC other positions (By similarity).
CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC methionyl peptide.
CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB008593; AAL02816.1; -.
DR PIR; F97734; F97734.
DR HAMAP; MF 00163; -.
DR InterPro; IPR000181; Pep_deformylase.
DR Pfam; PF01327; Pep_deformylase; 1.
DR PRINTS; PR01576; PDEFORMYLASE.
DR ProDom; PD003844; Pep_deformylase; 1.
DR TIGRfam; TIGR00079; pept_deformyl; 1.
KW Protein biosynthesis; Hydrolase; iron; Complete proteome.
FT ACT_SITE 142 142 BY SIMILARITY.
FT METAL 99 99 IRON (BY SIMILARITY).
FT METAL 141 141 IRON (BY SIMILARITY).
FT METAL 145 145 IRON (BY SIMILARITY).
SQ SEQUENCE 175 AA; 20279 MW; 93B050C6003C5A6D CRC64;

Query Match 66.0%; Score 35; DB 1; Length 175;
Best Local Similarity 62.5%; Pred. No. 14;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELI 8
Db 79 FIVNPEII 86

RESULT 13
DEF1_RICCN STANDARD; PRT; 175 AA.
ID DEF1_RICCN Q9ZDV8;
AC Q9ZDV8;
DT 30-MAY-2000 (Rel. 39, Created)
```

```
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).
GN DEF OR RP208.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Madrid E;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC newly synthesized proteins. Requires at least a dipeptide for an
CC efficient rate of reaction. N-terminal L-methionine is a
CC prerequisite for activity but the enzyme has broad specificity at
CC other positions (By similarity).
CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC methionyl peptide.
CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235270; CAJ14673.1; -.
DR PIR; B71732; B71732.
DR HSSP; P27251; 1DEF.
DR HAMAP; MF 00163; -.
DR InterPro; IPR000181; Pep_deformylase.
DR Pfam; PF01327; Pep_deformylase; 1.
DR PRINTS; PR01576; PDEFORMYLASE.
DR ProDom; PD003844; Pep_deformylase; 1.
DR TIGRfam; TIGR00079; pept_deformyl; 1.
KW Protein biosynthesis; Hydrolase; iron; Complete proteome.
FT ACT_SITE 142 142 BY SIMILARITY.
FT METAL 99 99 IRON (BY SIMILARITY).
FT METAL 141 141 IRON (BY SIMILARITY).
FT METAL 145 145 IRON (BY SIMILARITY).
SQ SEQUENCE 175 AA; 20269 MW; 2400553B491ABED9 CRC64;

Query Match 66.0%; Score 35; DB 1; Length 175;
Best Local Similarity 62.5%; Pred. No. 14;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELI 8
Db 79 FIVNPEII 86

RESULT 14
IF2G_YEAST STANDARD; PRT; 527 AA.
ID IF2G_YEAST P32481;
AC P32481;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2 gamma subunit (eIF-2-
DE gamma).
GN GCD11 OR TIF213 OR YER025W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```



OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93109337; PubMed=8417348;  
RA Hannig E.M., Cigan A.M., Freeman B.A., Kinzy T.G.;  
RT "GCD11, a negative regulator of GCN4 expression, encodes the gamma  
RT subunit of eIF-2 in Saccharomyces cerevisiae.";  
RL Mol. Cell. Biol. 13:506-520(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RX MEDLINE=97313264; PubMed=9169868;  
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,  
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,  
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,  
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,  
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,  
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,  
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";  
RL Nature 387:78-81(1997).  
CC -!- FUNCTION: eIF-2 functions in the early steps of protein synthesis  
CC by forming a ternary complex with GTP and initiator tRNA. This  
CC complex binds to a 40s ribosomal subunit, followed by mRNA binding  
CC to form a 43S preinitiation complex. Junction of the 60S ribosomal  
CC subunit to form the 80S initiation complex is preceded by  
CC hydrolysis of the GTP bound to eIF-2 and release of an eIF-2-GDP  
CC binary complex. In order for eIF-2 to recycle and catalyze another  
CC round of initiation, the GDP bound to eIF-2 must exchange with GTP  
CC by way of a reaction catalyzed by eIF-2b.  
CC -!- SUBUNIT: Heterotrimer composed of an alpha, a beta and a gamma  
CC chain.  
CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.  
CC EIF2G subfamily.  
CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L04268; AAA34633.1; -;  
CC EMBL; U18778; AAB64558.1; -;  
CC PIR; A48117; A48117.  
CC Germline; 139105; -;  
CC SGD; S0000827; GCD11.  
CC InterPro; IPR000795; EF\_GTPbind.  
CC InterPro; IPR004161; EFTU\_D2.  
CC InterPro; IPR009001; Elong\_init\_C.  
CC InterPro; IPR009000; Translat\_factor.  
CC Pfam; PF00009; GTP\_EFTU; 1.  
CC Pfam; PF03144; GTP\_EFTU\_D2; 1.  
CC PRINTS; PR00315; ELONGATNFCT.  
KW Initiation factor; Protein biosynthesis; GTP-binding.  
FT NP\_BIND 107 114 GTP (BY SIMILARITY).  
FT NP\_BIND 193 197 GTP (BY SIMILARITY).  
FT NP\_BIND 249 252 GTP (BY SIMILARITY).  
FT BINDING 194 194 ASSOCIATED WITH AMINOACYL-TRNA BINDING  
FT (BY SIMILARITY).  
FT DOMAIN 155 179 CYS-RICH.  
SQ SEQUENCE 527 AA; 57865 MW; D498AE62BC3E81CD CRC64;

Query Match 66.0%; Score 35; DB 1; Length 527;  
Best Local Similarity 63.6%; Pred. No. 46;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
Db 309 FMISPRLLIVR 319

RESULT 15  
Y422 MYCPN STANDARD; PRT; 839 AA.  
AC P75175;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MG422 homolog (C12\_orf839).  
GN MPN620 OR MF222.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97105885; PubMed=8948633;  
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
RA Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
RT pneumoniae.";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE000022; AAB95870.1; -;  
CC PIR; S73548; S73548.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 839 AA; 99912 MW; 3E3523E18BCDFECC CRC64;

Query Match 66.0%; Score 35; DB 1; Length 839;  
Best Local Similarity 66.7%; Pred. No. 76;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LINPELIVS 10  
Db 338 LLNPDLVLS 346

Search completed: May 26, 2004, 18:41:20  
Job time : 1.17624 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 0.877872 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 53    | 100.0       | 920    | 11 Q8CB72 | Q8cb72 mus musculu |
| 2          | 53    | 100.0       | 961    | 11 Q80YQ6 | Q80yq6 mus musculu |
| 3          | 53    | 100.0       | 968    | 11 Q88622 | Q88622 mus musculu |
| 4          | 53    | 100.0       | 972    | 11 Q9QYM2 | Q9qym2 rattus norv |
| 5          | 53    | 100.0       | 977    | 6 Q02776  | Q02776 bos taurus  |
| 6          | 52    | 98.1        | 976    | 4 Q9Y4W7  | Q9y4w7 homo sapien |
| 7          | 52    | 98.1        | 976    | 4 Q86W56  | Q86w56 homo sapien |
| 8          | 52    | 98.1        | 976    | 4 Q7Z742  | Q7z742 homo sapien |
| 9          | 42    | 79.2        | 364    | 10 Q9SKB4 | Q9skb4 arabidopsis |
| 10         | 41    | 77.4        | 249    | 16 Q9Z7P9 | Q9z7p9 chlamydia p |
| 11         | 39    | 73.6        | 312    | 2 Q8KLF8  | Q8klf8 rhizobium e |
| 12         | 38    | 71.7        | 150    | 16 Q895Q2 | Q895q2 clostridium |
| 13         | 38    | 71.7        | 386    | 16 Q82VH5 | Q82vh5 nitrosomona |
| 14         | 38    | 71.7        | 472    | 16 Q8F462 | Q8f462 leptospira  |
| 15         | 38    | 71.7        | 548    | 10 Q9SKB3 | Q9skb3 arabidopsis |
| 16         | 38    | 71.7        | 1063   | 5 Q8SVI4  | Q8svi4 encephalito |

|    |    |      |       |           |                    |
|----|----|------|-------|-----------|--------------------|
| 17 | 37 | 69.8 | 161   | 12 Q9DVM2 | Q9dvw2 plutella xy |
| 18 | 37 | 69.8 | 302   | 3 Q04110  | Q04110 saccharomyc |
| 19 | 37 | 69.8 | 522   | 10 Q8VYA1 | Q8vyal arabidopsis |
| 20 | 37 | 69.8 | 13414 | 5 Q8I6I6  | Q8i6i6 cryptospori |
| 21 | 36 | 67.9 | 185   | 16 Q82TC8 | Q82tc8 nitrosomona |
| 22 | 36 | 67.9 | 194   | 16 Q7V1K9 | Q7v1k9 prochloroco |
| 23 | 36 | 67.9 | 200   | 16 Q9K7M6 | Q9k7m6 bacillus ha |
| 24 | 36 | 67.9 | 219   | 16 Q9KRK2 | Q9krk2 vibrio chol |
| 25 | 36 | 67.9 | 219   | 16 Q87QD7 | Q87qd7 vibrio para |
| 26 | 36 | 67.9 | 230   | 2 Q9EYK4  | Q9eyk4 lactobacill |
| 27 | 36 | 67.9 | 255   | 16 Q88X77 | Q88x77 lactobacill |
| 28 | 36 | 67.9 | 314   | 16 Q92XV5 | Q92xv5 rhizobium m |
| 29 | 36 | 67.9 | 320   | 16 Q89AZ9 | Q89az9 buchnera ap |
| 30 | 36 | 67.9 | 519   | 10 Q48655 | Q48655 oryza sativ |
| 31 | 36 | 67.9 | 520   | 10 Q84S50 | Q84s50 oryza sativ |
| 32 | 36 | 67.9 | 536   | 10 Q943I5 | Q943i5 oryza sativ |
| 33 | 36 | 67.9 | 557   | 16 Q9X0V0 | Q9x0v0 thermotoga  |
| 34 | 36 | 67.9 | 688   | 16 Q8A213 | Q8a213 bacteroides |
| 35 | 35 | 66.0 | 160   | 16 Q9KKN4 | Q9kkn4 vibrio chol |
| 36 | 35 | 66.0 | 160   | 17 Q8TIS4 | Q8tis4 methanosarc |
| 37 | 35 | 66.0 | 197   | 16 Q7VBW4 | Q7vbw4 prochloroco |
| 38 | 35 | 66.0 | 208   | 16 Q97M04 | Q97m04 clostridium |
| 39 | 35 | 66.0 | 226   | 17 Q97X84 | Q97x84 sulfolobus  |
| 40 | 35 | 66.0 | 246   | 16 Q99T40 | Q99t40 staphylococ |
| 41 | 35 | 66.0 | 246   | 16 Q8NVV9 | Q8nvv9 staphylococ |
| 42 | 35 | 66.0 | 297   | 2 Q53036  | Q53036 haemophilus |
| 43 | 35 | 66.0 | 321   | 17 Q974E9 | Q974e9 sulfolobus  |
| 44 | 35 | 66.0 | 343   | 16 Q9RZB5 | Q9rzb5 deinococcus |
| 45 | 35 | 66.0 | 377   | 17 Q9YBM2 | Q9ybm2 aeropyrum p |

ALIGNMENTS

RESULT 1  
Q8CB72 ID Q8CB72 PRELIMINARY; PRT; 920 AA.  
AC Q8CB72;  
DT 01-MAR-2003 (TremBLrel. 23, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)  
DE Poly.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
DR EMBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 100.0%; Score 53; DB 11; Length 920;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
Db 752 FLINPELIVSR 762

RESULT 2  
Q80YQ6

```
ID Q80YQ6 PRELIMINARY; PRT; 961 AA.
AC Q80YQ6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Poly (ADP-ribose) glycohydrolase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC050892; AAH50892.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

Query Match 100.0%; Score 53; DB 11; Length 961;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
Db 752 FLINPELIVSR 762

RESULT 3
O88622 PRELIMINARY; PRT; 968 AA.
AC O88622;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly (ADP-ribose) glycohydrolase.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380098; PubMed=10449915;
RA Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT hybridization."
RL Cytogenet. Cell Genet. 85:269-270(1999).
DR EMBL; AF079557; AAC28735.1; -.
DR MGD; MGI:1347094; Parg.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;

Query Match 100.0%; Score 53; DB 11; Length 968;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
Db 751 FLINPELIVSR 761

RESULT 4
Q9QYM2 PRELIMINARY; PRT; 972 AA.
ID Q9QYM2
AC Q9QYM2;
```

```
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly (ADP-ribose) glycohydrolase.
GN PARG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BUF; TISSUE=Colon;
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,
RA Aoki Y., Nakagama H., Sugimura T.;
RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase
RT (Parg).";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AB019366; BAA87901.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;

Query Match 100.0%; Score 53; DB 11; Length 972;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
Db 755 FLINPELIVSR 765

RESULT 5
O02776 PRELIMINARY; PRT; 977 AA.
AC O02776;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly (ADP-ribose) glycohydrolase.
GN BPARG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97277328; PubMed=9115250;
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-
RT ribose) glycohydrolase."
RL J. Biol. Chem. 272:11895-11901(1997).
DR EMBL; U78975; AAB53370.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;

Query Match 100.0%; Score 53; DB 6; Length 977;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
Db 760 FLINPELIVSR 770

RESULT 6
Q9Y4W7 PRELIMINARY; PRT; 976 AA.
ID Q9Y4W7
```

AC Q9Y4W7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93380098; PubMed=10449915;  
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
RT hybridization.";  
RL Cytogenet. Cell Genet. 85:269-270(1999).  
DR EMBL; AF005043; AAB61614.1; -.  
DR Genew; HGNC:8605; PARG.  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;

Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.27;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
Db 759 FLINPELIISR 769

RESULT 7  
Q86W56  
ID Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;

Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.27;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
Db 759 FLINPELIISR 769

RESULT 8  
Q7Z742  
ID Q7Z742 PRELIMINARY; PRT; 976 AA.  
AC Q7Z742;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC052966; AAH52966.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;

Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.27;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
Db 759 FLINPELIISR 769

RESULT 9  
Q9SKB4  
ID Q9SKB4 PRELIMINARY; PRT; 364 AA.  
AC Q9SKB4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Putative poly(ADP-ribose) glycohydrolase.  
GN AT2G31860.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,



RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana.";  
RL Nature 402:761-768(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006533; AAD32286.1; -;  
DR PIR; A84726; A84726.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 364 AA; 40896 MW; 5DECA51A72089A94 CRC64;

Query Match 79.2%; Score 42; DB 10; Length 364;  
Best Local Similarity 88.9%; Pred. No. 8.8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIV 9  
|:|||||  
Db 202 FMINPELIV 210

RESULT 10  
Q9Z7P9 PRELIMINARY; PRT; 249 AA.  
AC Q9Z7P9;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE DNA polymerase III epsilon chain.  
GN DNAQ 2 OR CPN0655 OR CP0092 OR CPB0681.  
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TW-183;  
RA Geng M.M., Schuhmacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,  
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;

RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
other Chlamydia strains based on whole genome sequence analysis.";  
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AE001648; AAD18794.1; -;  
DR EMBL; AE002172; AAF73628.1; -;  
DR EMBL; AP002547; BAA98862.1; -;  
DR EMBL; AE017159; AAP98610.1; -;  
DR PIR; D86572; D86572.  
DR PIR; H72051; H72051.  
DR TIGR; CP0092; -;  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0004527; F:exonuclease activity; IEA.  
DR GO; GO:0006260; P:DNA replication; IEA.  
DR InterPro; IPR006054; DnaQ.  
DR InterPro; IPR006055; Exonuclease.  
DR Pfam; PF00929; Exonuclease; 1.  
DR SMART; SM00479; EXOIII; 1.  
DR TIGRFAMs; TIGR00573; dnaq; 1.  
KW Complete proteome.  
SQ SEQUENCE 249 AA; 28718 MW; E4B429A4BB895705 CRC64;

Query Match 77.4%; Score 41; DB 16; Length 249;  
Best Local Similarity 80.0%; Pred. No. 9.3;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIVS 10  
|||||:|  
Db 45 FLINPERVVS 54

RESULT 11  
Q8KLF8 PRELIMINARY; PRT; 312 AA.  
AC Q8KLF8;  
DT 01-OCT-2002 (TReMBLrel. 22, Created)  
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN NODDI.  
OS Rhizobium etli.  
OG Plasmid symbiotic plasmid p42d.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
OX NCBI\_TaxID=29449;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CFN42;  
RX MEDLINE=91193195; PubMed=2013564;  
RA Girard M.L., Flores M., Brom S., Romero D., Palacios R., Davila G.;  
RT "Structural complexity of the symbiotic plasmid of Rhizobium  
leguminosarum bv. phaseoli.";  
RL J. Bacteriol. 173:2411-2419(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CFN42;  
RX MEDLINE=97419521; PubMed=9274036;  
RA Ramirez-Romero M.A., Bustos P., Girard L., Rodriguez O.,  
RA Cevallos M.A., Davila G.;  
RT "Sequence, localization and characteristics of the replicator region  
of the symbiotic plasmid of Rhizobium etli.";  
RL Microbiology 143:2825-2831(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CFN42;  
RX Quintero V., Cevallos M.A., Davila G.;  
RT "A site-specific recombinase and RecA are required to exert  
incompatibility towards the symbiotic plasmid of Rhizobium etli.";  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
DR EMBL; U80928; AAM54780.1; -;  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.



OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,  
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,  
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,  
RA Fraser C.M., Venter J.C.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Panda S., Poirier G.G., Kay S.A.;  
RT "TEJ defines a role for poly-ADP-ribosylation in establishing period  
RT length of the Arabidopsis circadian oscillator.";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006533; AAD32285.2; -.  
DR EMBL; AF394690; AAK72256.1; -.  
DR PIR; B84726; B84726.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 548 AA; 62169 MW; F1A79FDA157C3329 CRC64;

Query Match 71.7%; Score 38; DB 10; Length 548;  
Best Local Similarity 87.5%; Pred. No. 82;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELI 8  
|:|||||  
Db 277 FMINPELI 284

Search completed: May 26, 2004, 18:46:12  
Job time : 2.87787 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 1.83313 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-14  
Perfect score: 71  
Sequence: 1 IALXLPNIXTQPIPLL 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 67    | 94.4        | 16     | 5     | Aae25637 Bovine po |
| 2          | 67    | 94.4        | 16     | 5     | Aau76028 Bovine po |
| 3          | 67    | 94.4        | 16     | 5     | Aau76019 Bovine po |
| 4          | 67    | 94.4        | 16     | 6     | Abg72286 Oligopept |
| 5          | 65    | 91.5        | 17     | 5     | Aae25649 Bovine po |
| 6          | 65    | 91.5        | 968    | 5     | Aae25631 Murine po |
| 7          | 65    | 91.5        | 968    | 5     | Aau76022 Mouse pol |
| 8          | 65    | 91.5        | 968    | 5     | Aau76013 Mouse pol |
| 9          | 65    | 91.5        | 968    | 6     | Abg72280 Mmurine p |
| 10         | 65    | 91.5        | 976    | 5     | Aae25630 Human pol |
| 11         | 65    | 91.5        | 976    | 5     | Aau76021 Human pol |
| 12         | 65    | 91.5        | 976    | 5     | Aau76012 Human pol |
| 13         | 65    | 91.5        | 976    | 6     | Abg72279 Human pol |
| 14         | 65    | 91.5        | 977    | 5     | Aae25629 Bovine po |
| 15         | 65    | 91.5        | 977    | 5     | Aau76020 Bovine po |
| 16         | 65    | 91.5        | 977    | 5     | Aau75799 Bovine po |
| 17         | 65    | 91.5        | 977    | 6     | Abg72278 Bovine po |
| 18         | 46    | 64.8        | 768    | 4     | Abb59491 Drosophil |
| 19         | 46    | 64.8        | 768    | 5     | Aae25632 Fruit fly |
| 20         | 46    | 64.8        | 768    | 5     | Aau76023 Fruit fly |
| 21         | 46    | 64.8        | 768    | 5     | Aau76014 Fruit fly |
| 22         | 46    | 64.8        | 768    | 6     | Abg72281 Fruit fly |
| 23         | 43    | 60.6        | 279    | 3     | Aag46848 Arabidops |
| 24         | 43    | 60.6        | 309    | 3     | Aag46847 Arabidops |
| 25         | 43    | 60.6        | 521    | 6     | Ada15481 A. thalia |

|    |    |      |     |   |          |           |
|----|----|------|-----|---|----------|-----------|
| 26 | 43 | 60.6 | 521 | 7 | ADD55676 | Thalecres |
| 27 | 43 | 60.6 | 521 | 7 | ADD30889 | Plant yie |
| 28 | 41 | 57.7 | 961 | 7 | ADB47689 | A. gossyp |
| 29 | 40 | 56.3 | 429 | 6 | ABU30155 | Protein e |
| 30 | 40 | 56.3 | 774 | 4 | ABB71927 | Drosophil |
| 31 | 39 | 54.9 | 109 | 6 | ABP76255 | Human GEN |
| 32 | 39 | 54.9 | 223 | 4 | ABG04588 | Novel hum |
| 33 | 39 | 54.9 | 228 | 7 | ADB82748 | Human pro |
| 34 | 39 | 54.9 | 335 | 4 | ABG09631 | Novel hum |
| 35 | 39 | 54.9 | 423 | 5 | AAO14067 | Human lys |
| 36 | 39 | 54.9 | 423 | 7 | ADE61363 | Human Pro |
| 37 | 39 | 54.9 | 423 | 7 | ADE61361 | Rat Prote |
| 38 | 39 | 54.9 | 821 | 4 | ABG25804 | Novel hum |
| 39 | 38 | 53.5 | 56  | 4 | AAM81219 | Human hae |
| 40 | 38 | 53.5 | 56  | 4 | AAM80769 | Human hae |
| 41 | 38 | 53.5 | 56  | 4 | AAM81836 | Human hae |
| 42 | 38 | 53.5 | 102 | 5 | ABP09179 | Human ORF |
| 43 | 38 | 53.5 | 220 | 3 | AAB21243 | DaEPV par |
| 44 | 38 | 53.5 | 329 | 5 | ABB47825 | Listeria  |
| 45 | 38 | 53.5 | 329 | 6 | ABU33015 | Protein e |

ALIGNMENTS

RESULT 1  
AAE25637  
ID AAE25637 standard; peptide; 16 AA.  
XX  
AC AAE25637;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #4.  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 4 /label= Unknown  
FT Misc-difference 9 /label= Unknown  
FT  
FT  
FT  
XX US6395543-B1.  
XX  
XX 28-MAY-2002.  
XX  
XX 23-FEB-2000; 2000US-00511507.  
XX  
XX 01-MAY-1998; 98US-0083768P.  
XX 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
XX ribose glycohydrolase involved in cellular response to DNA damage,  
XX inhibition of which is useful for treating neoplastic disorders and  
XX neurodegenerative diseases.  
XX  
XX Claim 10; Col 81-82; 77pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule which encodes

CC



CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is bPARG oligopeptide  
XX  
SQ

Sequence 16 AA;

Query Match 94.4%; Score 67; DB 5; Length 16;  
Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
|||||  
Db 1 IALXLPNIXTQPIPLL 16

RESULT 2

AAU76028  
ID AAU76028 standard; peptide; 16 AA.

XX AC AAU76028;

XX DT 08-MAY-2002 (first entry)

XX DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 75.

XX KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
XX adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
XX leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
XX hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
XX neurodegenerative disease; neurological disorder; Alzheimer's disease;  
XX Huntington's disease; Parkinson's disease; oligopeptide 75.

OS Bos taurus.

XX FH Key Location/Qualifiers

XX FT Misc-difference 4 /label= Unknown  
XX FT /note= "Xaa is not further defined in the specification"  
XX FT Misc-difference 9 /label= Unknown  
XX FT /note= "Xaa is not further defined in the specification"

XX PN US6337202-B1.

XX XX 08-JAN-2002.

XX PF 23-FEB-2000; 2000US-00511477.

XX PR 01-MAY-1998; 98US-0083768P.

XX PR 30-APR-1999; 99US-00302812.

XX XX (KENT ) UNIV KENTUCKY RES FOUND.

XX XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX XX WPI; 2002-163240/21.

XX XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.

XX XX Claim 4; Col 25; 81pp; English.

XX PS The present invention relates to a new poly(ADP-ribose) glycohydrolase

CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents bovine PARG  
CC oligonucleotide 75. This peptide is one of several PARG oligopeptides  
CC (AAU76025-AAU76028) of the invention  
XX

SQ Sequence 16 AA;

Query Match 94.4%; Score 67; DB 5; Length 16;  
Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
|||||  
Db 1 IALXLPNIXTQPIPLL 16

RESULT 3

AAU76019  
ID AAU76019 standard; peptide; 16 AA.

XX AC AAU76019;

XX DT 08-MAY-2002 (first entry)

XX DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 75.

XX KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
XX inherited genetic disease; myocardial infarction; vascular stroke; aging;  
XX neurodegeneration; Huntington's disease; Parkinson's disease;  
XX Alzheimer's disease; neurotoxicity; oligopeptide 75.

OS Bos taurus.

XX FH Key Location/Qualifiers

XX FT Misc-difference 4 /label= Unknown  
XX FT /note= "Xaa is not further defined in the specification"  
XX FT Misc-difference 9 /label= Unknown  
XX FT /note= "Xaa is not further defined in the specification"

XX PN US6333148-B1.

XX XX 25-DEC-2001.

XX PF 30-APR-1999; 99US-00302812.

XX PR 01-MAY-1998; 98US-0083768P.

XX XX (KENT ) UNIV KENTUCKY RES FOUND.

XX XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX XX WPI; 2002-153820/20.

XX XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.

XX XX Claim 4; Col 25; 80pp; English.

XX XX The present invention relates to a new method for screening compounds for

CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents bovine PARG oligopeptide 75. This peptide is one of several  
 CC PARG oligopeptides (AAU76016-AAU76019) of the invention

SQ Sequence 16 AA;  
 Query Match 94.4%; Score 67; DB 5; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16  
 Db 1 IALXLPNIXTQPIPLL 16

RESULT 4  
 ABG72286  
 ID ABG72286 standard; peptide; 16 AA.  
 XX  
 AC ABG72286;  
 XX  
 DT 13-MAR-2003 (first entry)  
 XX  
 DE Oligopeptide #4 derived from bovine PARG enzyme.  
 XX  
 KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; neurotropic; antiparkinsonian; cardiant; vasotrophic;  
 KW anticonvulsant; cerebroprotective.  
 XX  
 OS Bos taurus.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 4  
 FT Misc-difference 9 /label= Unknown  
 FT Misc-difference 9 /label= Unknown  
 FT  
 XX  
 PN US2002132328-A1.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 09-OCT-2001; 2001US-00973451.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEJ/) AME J.  
 PA (LINW/) LIN W.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2003-155895/15.  
 XX  
 PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX

PS Example 2; Page 14; 86pp; English.  
 XX  
 CC The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. ABG72283-  
 CC ABG72286 represent oligopeptides derived from bovine PARG enzyme. The  
 CC oligopeptides are used to construct degenerate PCR primers for the  
 CC isolation of cDNA encoding bovine PARG

XX  
 SQ Sequence 16 AA;  
 Query Match 94.4%; Score 67; DB 6; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16  
 Db 1 IALXLPNIXTQPIPLL 16

RESULT 5  
 AAE25649  
 ID AAE25649 standard; peptide; 17 AA.  
 XX  
 AC AAE25649;  
 XX  
 DT 04-NOV-2002 (first entry)  
 XX  
 DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #5.  
 XX  
 KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytostatic; vasotrophic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 OS Bos taurus.  
 XX  
 PN US6395543-B1.  
 XX  
 PD 28-MAY-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511507.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2002-535641/57.  
 XX  
 PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 PS Example 3; Col 27; 77pp; English.  
 XX

CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is bPARG oligopeptide

XX  
SQ Sequence 17 AA;

Query Match 91.5%; Score 65; DB 5; Length 17;  
Best Local Similarity 87.5%; Pred. No. 0.00023;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16  
Db 1 IALCLPNICTQPIPLL 16

RESULT 6  
AAE25631  
ID AAE25631 standard; protein; 968 AA.

XX  
AC AAE25631;  
XX  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
XX  
KW Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.

XX Mus musculus.

OS  
PN US6395543-B1.

XX  
PD 28-MAY-2002.

XX  
PF 23-FEB-2000; 2000US-00511507.

XX  
PR 01-MAY-1998; 98US-0083768P.

XX  
PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX N-PSDB; AAD42083.

XX  
PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.

XX Claim 3; Col 63-68; 77pp; English.

XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the

CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX  
SQ Sequence 968 AA;

Query Match 91.5%; Score 65; DB 5; Length 968;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16  
Db 592 IALCLPNICTQPIPLL 607

RESULT 7

AAU76022

ID AAU76022 standard; protein; 968 AA.

XX  
AC AAU76022;

XX  
DT 08-MAY-2002 (first entry)

XX  
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX  
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.

XX Mus musculus.

OS  
PN US6337202-B1.

XX  
PD 08-JAN-2002.

XX  
PF 23-FEB-2000; 2000US-00511477.

XX  
PR 01-MAY-1998; 98US-0083768P.

XX  
PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-163240/21.

XX N-PSDB; ABK14933.

XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 2; Col 63-70; 81pp; English.

XX  
CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the mouse PARG

CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX  
SQ Sequence 968 AA;

Query Match 91.5%; Score 65; DB 5; Length 968;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 592 IALCLPNICTQPIPLL 607

RESULT 8  
AAU76013  
ID AAU76013 standard; protein; 968 AA.  
XX  
AC AAU76013;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX  
OS Mus musculus.  
XX  
PN US6333148-B1.  
XX  
PD 25-DEC-2001.  
XX  
PF 30-APR-1999; 99US-00302812.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-153820/20.  
DR N-PSDB; ABK14495.  
XX  
PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
PS Claim 3; Col 63-68; 80pp; English.  
XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the mouse PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX  
SQ Sequence 968 AA;

Query Match 91.5%; Score 65; DB 5; Length 968;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 592 IALCLPNICTQPIPLL 607

RESULT 9  
ABG72280  
ID ABG72280 standard; protein; 968 AA.  
XX  
AC ABG72280;  
XX  
DT 13-MAR-2003 (first entry)  
XX  
DE Murine poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX  
KW Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
KW cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiant;  
KW vasotropic; anticonvulsant; cerebroprotective; enzyme.  
XX  
OS Mus musculus.  
XX  
PN US2002132328-A1.  
XX  
PD 19-SEP-2002.  
XX  
PF 09-OCT-2001; 2001US-00973451.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2003-155895/15.  
DR N-PSDB; ABX14479.  
XX  
PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
PS Claim 28; Fig 16; 86pp; English.  
XX  
CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC diseases (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents murine PARG enzyme  
XX  
SQ Sequence 968 AA;

Query Match 91.5%; Score 65; DB 6; Length 968;  
Best Local Similarity 87.5%; Pred. No. 0.018;



Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 592 IALCLPNICTQPIPLL 607

RESULT 10  
AAE25630  
ID AAE25630 standard; protein; 976 AA.  
XX AC AAE25630;  
DT 04-NOV-2002 (first entry)  
XX Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
DE Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
XX ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal injury; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
KW cytosstatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX Homo sapiens.  
OS US6395543-B1.  
XX 28-MAY-2002.  
XX 23-FEB-2000; 2000US-005111507.  
XX 01-MAY-1998; 98US-0083768P.  
XX 30-APR-1999; 99US-00302812.  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
DR N-PSDB; AAD42082.  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.  
XX Claim 3; Col 55-60; 77pp; English.  
XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The antisense therapeutics are useful to treat neoplastic disorders and conditions caused by genotoxic oxidative stress e.g., cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. The invention is useful in gene therapy and antisense therapy. The present sequence is human PARG

Sequence 976 AA;  
Query Match 91.5%; Score 65; DB 5; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 600 IALCLPNICTQPIPLL 615

RESULT 11  
AAU76021  
ID AAU76021 standard; protein; 976 AA.  
XX AC AAU76021;  
XX 08-MAY-2002 (first entry)  
XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
DE Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
XX adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX Homo sapiens.  
OS US6337202-B1.  
XX 08-JAN-2002.  
XX 23-FEB-2000; 2000US-005111477.  
XX 01-MAY-1998; 98US-0083768P.  
XX 30-APR-1999; 99US-00302812.  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-163240/21.  
XX N-PSDB; ABK14932.  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein which catalyses release of ADP-ribose from ADP ribose polymer, useful for treating neoplastic and neurological disorders, heart attack and stroke.  
XX Claim 2; Col 55-60; 81pp; English.  
XX The present invention relates to a new poly(ADP-ribose) glycohydrolase (PARG) protein which catalyses release of ADP-ribose from an ADP (adenosine diphosphate)-ribose polymer. The PARG molecule of the invention is useful for generating antibodies and can be inhibited or activated for diagnosing and treating neoplastic disorders such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, hyperplasia and hypertrophy, reperfusion following ischaemia, heart attack, stroke, neurodegenerative diseases, neurological disorders including Alzheimer's, Huntington's and Parkinson's diseases, and related conditions. PARG levels may be enhanced to suppress DNA repair and increase the cell's susceptibility to chemotherapy drugs. Antagonists of PARG are administered to treat or prevent neoplastic disorders. The present amino acid sequence represents the human PARG protein of the invention. This protein is one of several PARG proteins (AAU76020-AAU76024) of the invention

Sequence 976 AA;  
Query Match 91.5%; Score 65; DB 5; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 600 IALCLPNICTQPIPLL 615

RESULT 12  
AAU76012  
ID AAU76012 standard; protein; 976 AA.  
XX

AAU76012;  
08-MAY-2002 (first entry)  
Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose; adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer; inherited genetic disease; myocardial infarction; vascular stroke; aging; neurodegeneration; Huntington's disease; Parkinson's disease; Alzheimer's disease; neurotoxicity.  
Homo sapiens.  
US6333148-B1.  
25-DEC-2001.  
30-APR-1999; 99US-00302812.  
01-MAY-1998; 98US-0083768P.  
(KENT ) UNIV KENTUCKY RES FOUND.  
Jacobson MK, Jacobson EL, Ame J, Lin W;  
WPI; 2002-153820/20.  
N-PSDB; ABK14494.  
Screening compounds for modulation of poly(ADP-ribose) glycohydrolase, useful potentially for treating diseases associated with DNA damage, e.g. cancer.  
Claim 3; Col 55-60; 80pp; English.  
The present invention relates to a new method for screening compounds for ability to modulate activity of an enzyme that hydrolyses ADP (adenosine diphosphate)-ribose from an ADP-ribose polymer. The compounds are inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and are used to treat or prevent any condition associated with DNA damage e.g. neoplasia, inherited genetic diseases, myocardial infarction, vascular stroke, aging and neurodegeneration e.g. Huntington's, Parkinson's or Alzheimer's diseases, or neurotoxicity generally. Compounds identified by the new method are more effective than known inhibitors and have fewer side effects. The present amino acid sequence represents the human PARG protein of the invention. This protein is one of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the invention  
Sequence 976 AA;  
Query Match 91.5%; Score 65; DB 5; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTQPIPLL 16  
Db 600 IALCLPNICTQPIPLL 615  
RESULT 13  
ABG72279  
ID ABG72279 standard; protein; 976 AA.  
XX ABG72279;  
AC  
XX 13-MAR-2003 (first entry)  
DT  
XX Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
DE  
XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression; cellular response; DNA damage; neoplastic disorder inducing agent; oxidative stress; neoplastic disorder; myocardial infarction;  
KW  
KW

vascular stroke; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; Huntington's disease; inborn genetic error; reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic; anticonvulsant; cerebroprotective; enzyme.  
Homo sapiens.  
US2002132328-A1.  
19-SEP-2002.  
09-OCT-2001; 2001US-00973451.  
01-MAY-1998; 98US-0083768P.  
30-APR-1999; 99US-00302812.  
(JACO/) JACOBSON M K.  
(JACO/) JACOBSON E L.  
(AMEJ/) AME J.  
(LINW/) LIN W.  
Jacobson MK, Jacobson EL, Ame J, Lin W;  
WPI; 2003-155895/15.  
N-PSDB; ABX14478.  
New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose) glycohydrolase activity, for preventing, treating, or ameliorating a disease condition, e.g. neoplastic disorder, myocardial infarction or vascular stroke.  
Claim 28; Fig 16; 86pp; English.  
The present invention relates to the isolation of poly(ADP-ribose) glycohydrolase (PARG) from several species, and the polynucleotide sequences encoding them. Methods for inhibiting PARG expression or overexpressing PARG are also disclosed. PARG is involved in the cellular response to DNA damage, and is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. The polynucleotide sequences encoding PARG and PARG modulators are useful for preventing, treating, or ameliorating diseases such as neoplastic disorders, myocardial infarction, vascular stroke, neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's disease or Huntington's disease), inborn genetic errors, reperfusion following ischaemia, aging, and neurotoxicity. The polynucleotide sequences are also useful in gene therapy. The methods are useful for identifying an agent that modulates PARG activity, identifying a mutant PARG allele in an individual, or screening candidate molecules for PARG modulating activity. The present sequence represents human PARG enzyme  
Sequence 976 AA;  
Query Match 91.5%; Score 65; DB 6; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTQPIPLL 16  
Db 600 IALCLPNICTQPIPLL 615  
RESULT 14  
AAE25629  
ID AAE25629 standard; protein; 977 AA.  
XX  
AC AAE25629;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
XX Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW

KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.

OS Bos taurus.

XX US6395543-B1.

PN 28-MAY-2002.

XX 23-FEB-2000; 2000US-005111507.

XX 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

PI WPI; 2002-535641/57.

XX N-PSDB; AAD42081.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.

XX Claim 3; Col 47-45; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is bPARG

XX Sequence 977 AA;

Query Match 91.5%; Score 65; DB 5; Length 977;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 601 IALCLPNICTQPIPLL 616

RESULT 15  
AAU76020  
ID AAU76020 standard; protein; 977 AA.

XX AAU76020;

AC 08-MAY-2002 (first entry)

XX Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.

OS Bos taurus.

XX Key

FT Region

FT Region

FT Region

FT Region

FT Region

XX US6337202-B1.

XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-005111477.

XX 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

PI WPI; 2002-163240/21.

XX N-PSDB; ABK14931.

XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 2; Col 47-52; 81pp; English.

XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
PS (PARG) protein which catalyses release of ADP-ribose from an ADP  
XX (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the bovine PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention

XX Sequence 977 AA;

Query Match 91.5%; Score 65; DB 5; Length 977;  
Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 601 IALCLPNICTQPIPLL 616

Search completed: May 26, 2004, 18:40:13  
Job time : 2.83313 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 0.512697 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-14  
Perfect score: 71  
Sequence: 1 IALXLPNIXTQPIPLL 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match |        | DB ID                  | Description       |
|------------|-------|-------------|--------|------------------------|-------------------|
|            |       | Match       | Length |                        |                   |
| 1          | 67    | 94.4        | 16     | 4 US-09-302-812-14     | Sequence 14, Appl |
| 2          | 67    | 94.4        | 16     | 4 US-09-511-477-14     | Sequence 14, Appl |
| 3          | 67    | 94.4        | 16     | 4 US-09-511-507-14     | Sequence 14, Appl |
| 4          | 65    | 91.5        | 968    | 4 US-09-302-812-6      | Sequence 6, Appl  |
| 5          | 65    | 91.5        | 968    | 4 US-09-511-477-6      | Sequence 6, Appl  |
| 6          | 65    | 91.5        | 968    | 4 US-09-511-507-6      | Sequence 6, Appl  |
| 7          | 65    | 91.5        | 976    | 4 US-09-302-812-4      | Sequence 4, Appl  |
| 8          | 65    | 91.5        | 976    | 4 US-09-511-477-4      | Sequence 4, Appl  |
| 9          | 65    | 91.5        | 976    | 4 US-09-511-507-4      | Sequence 4, Appl  |
| 10         | 65    | 91.5        | 977    | 4 US-09-302-812-2      | Sequence 2, Appl  |
| 11         | 65    | 91.5        | 977    | 4 US-09-511-477-2      | Sequence 2, Appl  |
| 12         | 65    | 91.5        | 977    | 4 US-09-511-507-2      | Sequence 2, Appl  |
| 13         | 46    | 64.8        | 768    | 4 US-09-302-812-8      | Sequence 8, Appl  |
| 14         | 46    | 64.8        | 768    | 4 US-09-511-477-8      | Sequence 8, Appl  |
| 15         | 46    | 64.8        | 768    | 4 US-09-511-507-8      | Sequence 8, Appl  |
| 16         | 40    | 56.3        | 160    | 4 US-09-489-039A-11236 | Sequence 11236, A |
| 17         | 40    | 56.3        | 373    | 4 US-09-254-077A-12    | Sequence 12, Appl |
| 18         | 38    | 53.5        | 342    | 4 US-09-543-681A-5179  | Sequence 5179, Ap |
| 19         | 38    | 53.5        | 453    | 1 US-08-439-131A-5     | Sequence 5, Appl  |
| 20         | 38    | 53.5        | 453    | 1 US-08-440-674-4      | Sequence 4, Appl  |
| 21         | 38    | 53.5        | 453    | 4 US-08-879-337-6      | Sequence 6, Appl  |
| 22         | 38    | 53.5        | 582    | 4 US-09-543-681A-4556  | Sequence 4556, Ap |
| 23         | 37    | 52.1        | 181    | 4 US-09-540-236-3607   | Sequence 3607, Ap |
| 24         | 37    | 52.1        | 341    | 4 US-09-489-039A-8294  | Sequence 8294, Ap |
| 25         | 37    | 52.1        | 366    | 4 US-09-540-236-3007   | Sequence 3007, Ap |
| 26         | 37    | 52.1        | 415    | 4 US-09-543-681A-6746  | Sequence 6746, Ap |
| 27         | 37    | 52.1        | 437    | 4 US-09-489-039A-12063 | Sequence 12063, A |

|    |    |      |     |                        |                   |
|----|----|------|-----|------------------------|-------------------|
| 28 | 37 | 52.1 | 542 | 4 US-09-614-891-10     | Sequence 10, Appl |
| 29 | 37 | 52.1 | 584 | 4 US-09-489-039A-14137 | Sequence 14137, A |
| 30 | 37 | 52.1 | 597 | 1 US-08-399-696-102    | Sequence 102, App |
| 31 | 36 | 50.7 | 233 | 4 US-09-489-039A-12718 | Sequence 12718, A |
| 32 | 36 | 50.7 | 331 | 4 US-09-454-279-12     | Sequence 12, Appl |
| 33 | 36 | 50.7 | 341 | 4 US-09-543-681A-4713  | Sequence 4713, Ap |
| 34 | 36 | 50.7 | 445 | 4 US-09-252-991A-22769 | Sequence 22769, A |
| 35 | 36 | 50.7 | 566 | 4 US-09-252-991A-17972 | Sequence 17972, A |
| 36 | 35 | 49.3 | 101 | 4 US-09-345-236B-111   | Sequence 111, App |
| 37 | 35 | 49.3 | 425 | 4 US-09-634-955B-19    | Sequence 19, Appl |
| 38 | 35 | 49.3 | 466 | 4 US-09-489-039A-12781 | Sequence 12781, A |
| 39 | 35 | 49.3 | 583 | 4 US-09-107-532A-5678  | Sequence 5678, Ap |
| 40 | 35 | 49.3 | 689 | 4 US-09-252-991A-31790 | Sequence 31790, A |
| 41 | 34 | 47.9 | 16  | 3 US-09-179-558-12     | Sequence 12, Appl |
| 42 | 34 | 47.9 | 16  | 4 US-09-722-825-12     | Sequence 12, Appl |
| 43 | 34 | 47.9 | 16  | 4 US-09-722-487-12     | Sequence 12, Appl |
| 44 | 34 | 47.9 | 16  | 4 US-09-722-708-12     | Sequence 12, Appl |
| 45 | 34 | 47.9 | 25  | 1 US-08-190-130-1      | Sequence 1, Appl  |

ALIGNMENTS

RESULT 1  
US-09-302-812-14  
; Sequence 14, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 14  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-14

Query Match 94.4%; Score 67; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 4.8e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16  
| | | | | | | | | | | | | | | |  
Db 1 IALXLPNIXTQPIPLL 16

RESULT 2  
US-09-511-477-14  
; Sequence 14, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812



Thu May 27 09:55:59 2004

```
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 14
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-477-14

Query Match          94.4%; Score 67; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16
   |||||
Db 1 IALXLPNIXTQPIPLL 16

RESULT 3
US-09-511-507-14
; Sequence 14, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 14
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-507-14

Query Match          94.4%; Score 67; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16
   |||||
Db 1 IALXLPNIXTQPIPLL 16

RESULT 4
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
```

```
; ORGANISM: Mus musculus
; FEATURE:
US-09-302-812-6

Query Match          91.5%; Score 65; DB 4; Length 968;
Best Local Similarity 87.5%; Pred. No. 0.0095;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16
   |||||
Db 592 IALCLPNICTQPIPLL 607

RESULT 5
US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-477-6

Query Match          91.5%; Score 65; DB 4; Length 968;
Best Local Similarity 87.5%; Pred. No. 0.0095;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16
   |||||
Db 592 IALCLPNICTQPIPLL 607

RESULT 6
US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-507-6

Query Match          91.5%; Score 65; DB 4; Length 968;
```

Best Local Similarity 87.5%; Pred. No. 0.0095;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
| | | | | | | | | |  
Db 592 IALCLPNICTQPIPLL 607

RESULT 7  
US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

Query Match 91.5%; Score 65; DB 4; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.0095;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
| | | | | | | | | |  
Db 600 IALCLPNICTQPIPLL 615

RESULT 8  
US-09-511-477-4  
; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-477-4

Query Match 91.5%; Score 65; DB 4; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.0095;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
| | | | | | | | | |

Db 600 IALCLPNICTQPIPLL 615

RESULT 9  
US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-507-4

Query Match 91.5%; Score 65; DB 4; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.0095;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
| | | | | | | | | |  
Db 600 IALCLPNICTQPIPLL 615

RESULT 10  
US-09-302-812-2  
; Sequence 2, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-2

Query Match 91.5%; Score 65; DB 4; Length 977;  
Best Local Similarity 87.5%; Pred. No. 0.0096;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
| | | | | | | | | |  
Db 601 IALCLPNICTQPIPLL 616

RESULT 11  
US-09-511-477-2

```

; Sequence 2, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-477-2

```

```

Query Match          91.5%; Score 65; DB 4; Length 977;
Best Local Similarity 87.5%; Pred. No. 0.0096;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 IALXLPNIXTQPIPLL 16
      ||| |||| |||| |||| ||||
Db      601 IALCLPNICTQPIPLL 616

```

```

RESULT 12
US-09-511-507-2
; Sequence 2, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-507-2

```

```

Query Match          91.5%; Score 65; DB 4; Length 977;
Best Local Similarity 87.5%; Pred. No. 0.0096;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 IALXLPNIXTQPIPLL 16
      ||| |||| |||| |||| ||||
Db      601 IALCLPNICTQPIPLL 616

```

```

RESULT 13
US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.

```

```

; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN2
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-302-812-8

```

```

Query Match          64.8%; Score 46; DB 4; Length 768;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 IALXLPNIXTQPIPLL 16
      :|| ||||: ||||
Db      219 LALRLPDLIQSPVPLL 234

```

```

RESULT 14
US-09-511-477-8
; Sequence 8, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-477-8

```

```

Query Match          64.8%; Score 46; DB 4; Length 768;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 IALXLPNIXTQPIPLL 16
      :|| ||||: ||||
Db      219 LALRLPDLIQSPVPLL 234

```

```

RESULT 15
US-09-511-507-8
; Sequence 8, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH

```

FILE REFERENCE: NIAD 201  
CURRENT APPLICATION NUMBER: US/09/511,507  
CURRENT FILING DATE: 2000-02-23  
PRIOR APPLICATION NUMBER: 09/302,812  
PRIOR FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 38  
SEQ ID NO 8  
LENGTH: 768  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
FEATURE:  
US-09-511-507-8

Query Match 64.8%; Score 46; DB 4; Length 768;  
Best Local Similarity 50.0%; Pred. No. 9.8;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16  
:|||||:|:|  
Db 219 LALRLPDLIQSPVPLL 234

Search completed: May 26, 2004, 18:49:32  
Job time : 0.512697 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 1.33495 Seconds  
(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-14  
Perfect score: 71  
Sequence: 1 IALXLPNIXTQPIPLL 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues  
Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 67    | 94.4        | 16     | 9  | US-09-973-451-14     |
| 2          | 65    | 91.5        | 968    | 9  | US-09-973-451-6      |
| 3          | 65    | 91.5        | 976    | 9  | US-09-973-451-4      |
| 4          | 65    | 91.5        | 977    | 9  | US-09-973-451-2      |
| 5          | 46    | 64.8        | 768    | 9  | US-09-973-451-8      |
| 6          | 43    | 60.6        | 521    | 12 | US-10-412-699B-82    |
| 7          | 43    | 60.6        | 521    | 12 | US-10-225-066A-922   |
| 8          | 43    | 60.6        | 521    | 14 | US-10-278-173-36     |
| 9          | 43    | 60.6        | 521    | 14 | US-10-278-536-34     |
| 10         | 43    | 60.6        | 521    | 15 | US-10-374-780A-2244  |
| 11         | 41    | 57.7        | 79     | 12 | US-10-424-599-182136 |
| 12         | 41    | 57.7        | 499    | 14 | US-10-156-761-9332   |
| 13         | 41    | 57.7        | 1089   | 15 | US-10-369-493-2154   |
| 14         | 41    | 57.7        | 1579   | 9  | US-09-801-368-368    |
| 15         | 41    | 57.7        | 1579   | 15 | US-10-369-493-2000   |

|    |    |      |     |    |                      |                   |
|----|----|------|-----|----|----------------------|-------------------|
| 16 | 40 | 56.3 | 429 | 12 | US-10-282-122A-58079 | Sequence 58079, A |
| 17 | 39 | 54.9 | 228 | 15 | US-10-012-697-1529   | GENERAL INFORMATI |
| 18 | 38 | 53.5 | 56  | 9  | US-09-796-692-1133   | Sequence 1133, Ap |
| 19 | 38 | 53.5 | 56  | 9  | US-09-796-692-1583   | Sequence 1583, Ap |
| 20 | 38 | 53.5 | 56  | 9  | US-09-796-692-2200   | Sequence 2200, Ap |
| 21 | 38 | 53.5 | 56  | 14 | US-10-040-862-1133   | Sequence 1133, Ap |
| 22 | 38 | 53.5 | 56  | 14 | US-10-040-862-1583   | Sequence 1583, Ap |
| 23 | 38 | 53.5 | 56  | 14 | US-10-040-862-2200   | Sequence 2200, Ap |
| 24 | 38 | 53.5 | 56  | 15 | US-10-057-475B-1133  | Sequence 1133, Ap |
| 25 | 38 | 53.5 | 56  | 15 | US-10-057-475B-1583  | Sequence 1583, Ap |
| 26 | 38 | 53.5 | 56  | 15 | US-10-057-475B-2200  | Sequence 2200, Ap |
| 27 | 38 | 53.5 | 56  | 15 | US-10-154-884B-1133  | Sequence 1133, Ap |
| 28 | 38 | 53.5 | 56  | 15 | US-10-154-884B-1583  | Sequence 1583, Ap |
| 29 | 38 | 53.5 | 56  | 15 | US-10-154-884B-2200  | Sequence 2200, Ap |
| 30 | 38 | 53.5 | 329 | 12 | US-10-282-122A-60939 | Sequence 60939, A |
| 31 | 38 | 53.5 | 348 | 9  | US-09-730-617-2      | Sequence 2, Appli |
| 32 | 38 | 53.5 | 426 | 12 | US-10-282-122A-61195 | Sequence 61195, A |
| 33 | 38 | 53.5 | 430 | 12 | US-10-282-122A-55560 | Sequence 55560, A |
| 34 | 38 | 53.5 | 453 | 8  | US-08-879-337-6      | Sequence 6, Appli |
| 35 | 38 | 53.5 | 453 | 15 | US-10-369-493-2273   | Sequence 2273, Ap |
| 36 | 38 | 53.5 | 461 | 15 | US-10-108-260A-3238  | Sequence 3238, Ap |
| 37 | 38 | 53.5 | 477 | 12 | US-10-425-114-65173  | Sequence 65173, A |
| 38 | 38 | 53.5 | 616 | 9  | US-09-925-300-1519   | Sequence 1519, Ap |
| 39 | 38 | 53.5 | 651 | 9  | US-09-815-242-11022  | Sequence 11022, A |
| 40 | 38 | 53.5 | 651 | 12 | US-10-282-122A-58169 | Sequence 58169, A |
| 41 | 38 | 53.5 | 702 | 12 | US-10-282-122A-47346 | Sequence 47346, A |
| 42 | 38 | 53.5 | 713 | 15 | US-10-369-493-13792  | Sequence 13792, A |
| 43 | 38 | 53.5 | 766 | 9  | US-09-801-368-316    | Sequence 316, App |
| 44 | 38 | 53.5 | 950 | 15 | US-10-369-493-6705   | Sequence 6705, Ap |
| 45 | 38 | 53.5 | 952 | 15 | US-10-369-493-6704   | Sequence 6704, Ap |

ALIGNMENTS

RESULT 1  
US-09-973-451-14  
; Sequence 14, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine J.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 14  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-14  
Query Match 94.4%; Score 67; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.00028;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16  
| | | | | | | | | | | | | | | |  
Db 1 IALXLPNIXTQPIPLL 16

RESULT 2

US-09-973-451-2  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

Query Match 91.5%; Score 65; DB 9; Length 977;  
Best Local Similarity 87.5%; Pred. No. 0.046;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTQPIPLL 16  
Db 601 IALCLPNICTQPIPLL 616

RESULT 5  
US-09-973-451-8  
; Sequence 8, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-973-451-8

Query Match 64.8%; Score 46; DB 9; Length 768;  
Best Local Similarity 50.0%; Pred. No. 43;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTQPIPLL 16  
Db 219 IALRLPDLIQSPVPLL 234

US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

Query Match 91.5%; Score 65; DB 9; Length 968;  
Best Local Similarity 87.5%; Pred. No. 0.046;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTQPIPLL 16  
Db 592 IALCLPNICTQPIPLL 607

RESULT 3  
US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4

Query Match 91.5%; Score 65; DB 9; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.046;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTQPIPLL 16  
Db 600 IALCLPNICTQPIPLL 615

```
RESULT 6
US-10-412-699B-82
; Sequence 82, Application US/10412699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Zhang, James
; APPLICANT: Fromm, Michael E.
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc J.
; APPLICANT: Pineda, Pierre E.
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddle, James S.
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Samaha, Raymond R.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Creelman, Robert A.
; APPLICANT: DuBell, Arnold N.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Kumimoto, Roderick
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI-0048CIP
; CURRENT APPLICATION NUMBER: US/10/412,699B
; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 82
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G176
US-10-412-699B-82

Query Match 60.6%; Score 43; DB 12; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches 8; Conservative 3; Mismatches 3; Indels 3; Gaps 0;

QY 3 LXLPNIXTQPIPLL 16
| | | | | | | | | |
Db 89 LFLPSMVTQPLPQL 102

RESULT 7
US-10-225-066A-922
; Sequence 922, Application US/10225066A
; Publication No. US20030226173A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: RATCLIFFE, Oliver
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: ADAM, Luc J
; APPLICANT: DUBELL, Arnold T
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: PILGRIM, Marsha L
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: REUBER, T. Lynne
; APPLICANT: CREELMAN, Robert A
; APPLICANT: PINEDA, Omaira
; APPLICANT: YU, Guo-Liang
; APPLICANT: BROUN, Pierre E
; TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI0036-2 US
; CURRENT APPLICATION NUMBER: US/10/225,066A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 09/837,444
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 1122
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 922
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-225-066A-922

Query Match 60.6%; Score 43; DB 12; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches 8; Conservative 3; Mismatches 3; Indels 3; Gaps 0;

QY 3 LXLPNIXTQPIPLL 16
| | | | | | | | | |
Db 89 LFLPSMVTQPLPQL 102

RESULT 8
US-10-278-173-36
; Sequence 36, Application US/10278173
; Publication No. US20030061637A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Pineda, Omaira
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Keddle, James
; APPLICANT: Heard, Jacqueline
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
; FILE REFERENCE: MBI-009
; CURRENT APPLICATION NUMBER: US/10/278,173
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US/09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: PatentIn Ver. 2.1
```

```

; SEQ ID NO 36
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G176
US-10-278-173-36

Query Match          60.6%; Score 43; DB 14; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches      8; Conservative 3; Mismatches 3; Indels 3; Gaps 0;

QY      3 LXLPNIXTQPIPLL 16
      | | | | | | | | | |
Db      89 LFLPSMVTQPLPQL 102

RESULT 9
US-10-278-536-34
; Sequence 34, Application US/10278536
; Publication No. US20030131386A1
; GENERAL INFORMATION:
; APPLICANT: Samaha, Raymond
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Keddie, James
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-011
; CURRENT APPLICATION NUMBER: US/10/278,536
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G176
US-10-278-536-34

Query Match          60.6%; Score 43; DB 14; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches      8; Conservative 3; Mismatches 3; Indels 3; Gaps 0;

QY      3 LXLPNIXTQPIPLL 16
      | | | | | | | | | |
Db      89 LFLPSMVTQPLPQL 102

RESULT 10
US-10-374-780A-2244
; Sequence 2244, Application US/10374780A
; Publication No. US20040019927A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne

US-10-424-599-182136
; Sequence 182136, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(532223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 182136
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(79)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: PAT_MRT3847_135482C.1.pep
US-10-424-599-182136

Query Match          60.6%; Score 43; DB 15; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches      8; Conservative 3; Mismatches 3; Indels 3; Gaps 0;

QY      3 LXLPNIXTQPIPLL 16
      | | | | | | | | | |
Db      89 LFLPSMVTQPLPQL 102

RESULT 11
US-10-424-599-182136
; Sequence 182136, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(532223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 182136
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(79)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: PAT_MRT3847_135482C.1.pep
US-10-424-599-182136
```



Query Match 57.7%; Score 41; DB 12; Length 79;  
Best Local Similarity 63.6%; Pred. No. 25;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 PNIXTQPIPLL 16  
| : : : :  
Db 61 PTITSQPLPLL 71

RESULT 12  
US-10-156-761-9332  
; Sequence 9332, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9332  
; LENGTH: 499  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9332

Query Match 57.7%; Score 41; DB 14; Length 499;  
Best Local Similarity 43.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
: : : : :  
Db 185 VAIGVPVLLTQSPVL 200

RESULT 13  
US-10-369-493-2154  
; Sequence 2154, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2154  
; LENGTH: 1089  
; TYPE: PRT  
; ORGANISM: Schizosaccharomyces pombe  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1089)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-2154

Query Match 57.7%; Score 41; DB 15; Length 1089;  
Best Local Similarity 66.7%; Pred. No. 4e+02;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPLL 16  
| : : : :  
Db 498 LPSIGKPIPLL 509

RESULT 14  
US-09-801-368-368  
; Sequence 368, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250Alman, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 368  
; LENGTH: 1579  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-368

Query Match 57.7%; Score 41; DB 9; Length 1579;  
Best Local Similarity 54.5%; Pred. No. 5.9e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPL 15  
| : : : :  
Db 132 LPKLSTQPVV 142

RESULT 15  
US-10-369-493-2000  
; Sequence 2000, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2000  
; LENGTH: 1579

```

; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-2000

Query Match      57.7%  Score 41;  DB 15;  Length 1579;
Best Local Similarity 54.5%;  Pred. No. 5.9e+02;
Matches 6;  Conservative 3;  Mismatches 2;  Indels 0;  Gaps 0;

Qy      5 LPNIXTQPIPL 15
      ||:||||:
Db      132 LPKLSTQVPV 142

Search completed: May 26, 2004, 19:19:03
Job time : 2.33495 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 0.440145 Seconds  
(without alignments)  
3496.721 Million cell updates/sec

Title: US-09-302-812-14  
Perfect score: 71  
Sequence: 1 IALXLPNIXTQPIPL 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 44    | 62.0        | 317    | 2 T31897 | hypothetical prote |
| 2          | 43    | 60.6        | 309    | 2 T48481 | SPF1-like protein  |
| 3          | 43    | 60.6        | 522    | 2 T31893 | hypothetical prote |
| 4          | 43    | 60.6        | 1256   | 2 AB2042 | hypothetical prote |
| 5          | 42    | 59.2        | 218    | 2 S29712 | segmentation prote |
| 6          | 42    | 59.2        | 432    | 2 E82577 | seryl-trNA synthet |
| 7          | 41    | 57.7        | 124    | 2 T26799 | hypothetical prote |
| 8          | 41    | 57.7        | 1072   | 2 T37742 | serine threonine-p |
| 9          | 41    | 57.7        | 1579   | 2 S59801 | protein kinase SSK |
| 10         | 41    | 57.7        | 3157   | 2 B70969 | probable PPE prote |
| 11         | 40    | 56.3        | 258    | 2 S35276 | probable export pr |
| 12         | 40    | 56.3        | 373    | 2 A56392 | beta-galactoside a |
| 13         | 40    | 56.3        | 429    | 2 I64048 | serine-trNA ligase |
| 14         | 39    | 54.9        | 421    | 2 S14742 | acid phosphatase ( |
| 15         | 39    | 54.9        | 423    | 1 S06167 | acid phosphatase ( |
| 16         | 39    | 54.9        | 423    | 2 A33395 | acid phosphatase ( |
| 17         | 39    | 54.9        | 526    | 2 T48467 | aspartyl aminopept |
| 18         | 38    | 53.5        | 329    | 2 AG1130 | penicillin acylase |
| 19         | 38    | 53.5        | 333    | 2 AF2394 | hypothetical prote |
| 20         | 38    | 53.5        | 416    | 2 S33473 | interleukin-1 rece |
| 21         | 38    | 53.5        | 435    | 2 T46443 | hypothetical prote |
| 22         | 38    | 53.5        | 453    | 2 A43765 | stsl+ protein - fi |
| 23         | 38    | 53.5        | 556    | 2 G81798 | long-chain-fatty-a |
| 24         | 38    | 53.5        | 556    | 2 H81068 | long-chain-fatty-a |
| 25         | 38    | 53.5        | 562    | 2 AB0253 | long-chain-fatty-a |
| 26         | 38    | 53.5        | 651    | 2 G64068 | DNA topoisomerase  |
| 27         | 38    | 53.5        | 706    | 2 JQ2210 | probable nucleic a |
| 28         | 38    | 53.5        | 766    | 1 S61694 | flocculation suppr |
| 29         | 38    | 53.5        | 950    | 2 T28793 | diacylglycerol kin |

|    |      |      |          |                    |
|----|------|------|----------|--------------------|
| 30 | 53.5 | 952  | 2 T28792 | diacylglycerol kin |
| 31 | 53.5 | 1087 | 2 S58147 | protein kinase - f |
| 32 | 53.5 | 1239 | 2 T42020 | class IV chitin sy |
| 33 | 53.5 | 1504 | 2 T17426 | FK506 polyketide s |
| 34 | 52.1 | 84   | 2 A86689 | prophage ps2 prote |
| 35 | 52.1 | 245  | 2 B36869 | probable export pr |
| 36 | 52.1 | 245  | 2 G90964 | flagellar biosynth |
| 37 | 52.1 | 245  | 2 G85812 | flagellar biosynth |
| 38 | 52.1 | 253  | 1 A40582 | type IV prepilin p |
| 39 | 52.1 | 253  | 2 D82273 | leader peptidase T |
| 40 | 52.1 | 285  | 2 H84219 | hypothetical prote |
| 41 | 52.1 | 430  | 1 YSEC   | serine-trNA ligase |
| 42 | 52.1 | 430  | 2 B90751 | serine trNA synthe |
| 43 | 52.1 | 430  | 2 H85614 | hypothetical prote |
| 44 | 52.1 | 430  | 2 AI0611 | seryl-trNA synthet |
| 45 | 52.1 | 450  | 2 AC1079 | PTS system, cellob |

ALIGNMENTS

RESULT 1

T31897 hypothetical protein C03A7.13 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T31897

R;Greco, T.; Bradshaw, H.; Elliott, G.

submitted to the EMBL Data Library, July 1997

A;Description: The sequence of C. elegans cosmid C03A7.

A;Reference number: Z21096

A;Accession: T31897

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-317 <GRE>

A;Cross-references: EMBL:AF016451; PIDN:AAB65994.1; GSPDB:GN00023; CESP:C03A7.13

A;Experimental source: strain Bristol N2; clone C03A7

C;Genetics:

A;Gene: CESP:C03A7.13

A;Map position: 5

A;Introns: 53/3; 119/1; 186/1

Query Match 62.0%; Score 44; DB 2; Length 317;  
Best Local Similarity 46.7%; Pred. No. 2.7;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPL 15

Db 173 VTMGLPNIASQDVPL 187

RESULT 2

T48481

SPF1-like protein - Arabidopsis thaliana

N;Alternate names: protein T28J14.40

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C;Accession: T48481

R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Len

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24493

A;Accession: T48481

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-309 <BEV>

A;Cross-references: EMBL:AL163652

A;Experimental source: cultivar Columbia; BAC clone T28J14

C;Genetics:

A;Map position: 5

A;Introns: 51/3; 93/3; 211/2; 254/2

A;Note: T28J14.40

Query Match 60.6%; Score 43; DB 2; Length 309;

Best Local Similarity 57.1%; Pred. No. 3.9; Mismatches 3; Indels 0; Gaps 0;  
Matches 8; Conservative 3;  
QY 3 LXLNIXTQPIPLL 16  
Db 89 LFLPSMVTQPLPQL 102

RESULT 3  
T31893  
hypothetical protein C03A7.11 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 17-Mar-2000  
C;Accession: T31893  
R;Greco, T.; Bradshaw, H.; Elliott, G.  
submitted to the EMBL Data Library, July 1997  
A;Description: The sequence of C. elegans cosmid C03A7.  
A;Reference number: Z21096  
A;Accession: T31893  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-522 <GRE>  
A;Cross-references: EMBL:AF016451; PIDN:AAB66005.1; GSPDB:GN00023; CESP:C03A7.11  
A;Experimental source: strain Bristol N2; clone C03A7  
C;Genetics:  
A;Gene: CESP:C03A7.11  
A;Map position: 5  
A;Introns: 53/3; 117/1; 184/1; 357/1; 399/1  
C;Superfamily: glucuronosyltransferase

Query Match 60.6%; Score 43; DB 2; Length 522;  
Best Local Similarity 46.7%; Pred. No. 7;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 15  
Db 171 VTIGLPNIASQWVPL 185

RESULT 4  
AB2042  
hypothetical protein all1888 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AB2042  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AB2042  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1256 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA073587.1; PID:g17130978; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all1888

Query Match 60.6%; Score 43; DB 2; Length 1256;  
Best Local Similarity 66.7%; Pred. No. 19;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPLL 16  
Db 333 LDNVISQPIPLL 344

RESULT 5  
S29712  
segmentation protein hairy - red flour beetle  
C;Species: Tribolium castaneum (red flour beetle)

C;Date: 19-Mar-1997 #sequence\_revision 01-Aug-1997 #text\_change 17-Mar-1999  
C;Accession: S29712  
R;Sommer, R.J.; Tautz, D.  
Nature 361, 448-450, 1993  
A;Title: Involvement of an orthologue of the Drosophila pair-rule gene hairy in segment  
A;Reference number: S29712; MUID:93156810; PMID:8429884  
A;Accession: S29712  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-218 <SOM>

Query Match 59.2%; Score 42; DB 2; Length 218;  
Best Local Similarity 56.2%; Pred. No. 4;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 148 IALVLPQTQASPLPLL 163

RESULT 6  
E82577  
seryl-tRNA synthetase XF2286 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C;Accession: E82577  
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: A82515; MUID:20365717; PMID:10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: E82577  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-432 <SIM>  
A;Cross-references: GB:AE004040; GB:AE003849; NID:g9107437; PIDN:AAF85085.1; GSPDB:GN001  
A;Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsuchako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF2286  
C;Superfamily: serine-tRNA ligase

Query Match 59.2%; Score 42; DB 2; Length 432;  
Best Local Similarity 46.7%; Pred. No. 8.6;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 15  
Db 106 VALGIPNLPPQDVPL 120

RESULT 7  
T26799  
hypothetical protein Y40H7A.11 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T26799  
R;McMurray, A.  
submitted to the EMBL Data Library, November 1998  
A;Reference number: Z20268



F;101-117/Domain: transmembrane #status predicted <TM2>

F;202-218/Domain: transmembrane #status predicted <TM3>  
F;236-252/Domain: transmembrane #status predicted <TM4>

Query Match 56.3%; Score 40; DB 2; Length 258;  
Best Local Similarity 70.0%; Pred. No. 11;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPL 14  
||| |||||:  
Db 36 LPGIVTQPLP 45

RESULT 12

A56392  
beta-galactoside alphas,2-fucosyltransferase I - rabbit  
C;Species: Oryctolagus cuniculus (domestic rabbit)  
C;Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 05-Nov-1999  
C;Accession: A56392  
R;Hitoshi, S.; Kusunoki, S.; Kanazawa, I.; Tsuji, S.  
J. Biol. Chem. 270, 8844-8850, 1995  
A;Title: Molecular cloning and expression of two types of rabbit beta-galactoside alphas  
A;Reference number: A56392; MUID:95238380; PMID:7721792  
A;Accession: A56392  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-373 <HIT>  
A;Cross-references: GB:X80226; NID:g854354; PIDN:CAA56513.1; PID:g854355  
C;Keywords: transmembrane protein

Query Match 56.3%; Score 40; DB 2; Length 373;  
Best Local Similarity 50.0%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 14  
: || | | | : || | |  
Db 37 LALSLPCLERQVPV 50

RESULT 13

I64048  
serine-tRNA ligase (EC 6.1.1.11) - Haemophilus influenzae (strain Rd KW20)  
N;Alternate names: seryl-tRNA synthetase  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002  
C;Accession: I64048  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630; PMID:7542800  
A;Accession: I64048  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-429 <TIGR>  
A;Cross-references: GB:U32696; GB:L42023; NID:g1573057; PIDN:AAC21787.1; PID:g1573062; T

C;Genetics:  
A;Gene: sers  
C;Function:  
A;Description: charges tRNA(Ser) with serine  
A;Pathway: protein biosynthesis  
C;Superfamily: serine-tRNA ligase  
C;Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 56.3%; Score 40; DB 2; Length 429;  
Best Local Similarity 46.7%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 15  
||| : ||| : |||  
Db 100 IALSIPNLPADEVPL 114

RESULT 14

S14742  
acid phosphatase (EC 3.1.3.2) precursor - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 31-Mar-2000  
C;Accession: S14742  
R;Geier, C.; von Figura, K.; Pohlmann, R.  
Biol. Chem. Hoppe-Seyler 372, 301-304, 1991  
A;Title: Molecular cloning of the mouse lysosomal acid phosphatase.  
A;Reference number: S14742; MUID:91282986; PMID:2059337  
A;Accession: S14742  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-421 <GEI>  
A;Cross-references: EMBL:X57199; NID:g52870; PIDN:CAA40485.1; PID:g52871  
C;Superfamily: mammalian acid phosphatase  
C;Keywords: lysosome; phosphoric monoester hydrolase

Query Match 54.9%; Score 39; DB 2; Length 421;  
Best Local Similarity 70.0%; Pred. No. 29;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 PNIXTQPIPL 15  
||| |||||:  
Db 130 PNISWQPIPV 139

RESULT 15

S06167  
acid phosphatase (EC 3.1.3.2) ACP2 precursor [validated] - human  
N;Alternate names: acid phosphatase, lysosomal  
C;Species: Homo sapiens (man)  
C;Date: 04-Dec-1992 #sequence\_revision 01-Dec-1995 #text\_change 08-Dec-2000  
C;Accession: S06167; S05525; S01155  
R;von Figura, K.  
submitted to the EMBL Data Library, June 1989  
A;Reference number: S06167  
A;Accession: S06167  
A;Molecule type: DNA  
A;Residues: 1-423 <VON>  
A;Cross-references: EMBL:X15525; NID:g34239; PIDN:CAA33542.1; PID:g1199524  
R;Geier, C.; von Figura, K.; Pohlmann, R.  
Eur. J. Biochem. 183, 611-616, 1989  
A;Title: Structure of the human lysosomal acid phosphatase gene.  
A;Reference number: S05525; MUID:89377828; PMID:2776754  
A;Accession: S05525  
A;Molecule type: DNA  
A;Residues: 1-29 <GEI>  
R;Pohlmann, R.; Krentler, C.; Schmidt, B.; Schroeder, W.; Lorkowski, G.; Culley, J.; Mer  
EMBO J. 7, 2343-2350, 1988  
A;Title: Human lysosomal acid phosphatase: cloning, expression and chromosomal assignment  
A;Reference number: S01155; MUID:89052645; PMID:3191910  
A;Accession: S01155  
A;Molecule type: mRNA  
A;Residues: 1-423 <POH>  
A;Cross-references: EMBL:X12548; NID:g34262; PIDN:CAA31064.1; PID:g34263  
A;Note: part of this sequence, including the amino end of the mature protein, was confir

C;Genetics:  
A;Gene: GDB:ACP2  
A;Cross-references: GDB:118963; OMIM:171650  
A;Map position: 11p11.2-11p11.11  
A;Introns: 38/3; 70/3; 99/3; 150/3; 183/3; 213/3; 258/1; 285/3; 321/2; 380/1  
C;Function:  
A;Description: catalyzes the hydrolysis of a wide range of phosphate esters  
C;Superfamily: mammalian acid phosphatase  
C;Keywords: glycoprotein; phosphohistidine; phosphoprotein; phosphoric monoester hydrolase  
F;1-30/Domain: signal sequence #status predicted <SIG>  
F;31-423/Product: acid phosphatase ACP2 #status experimental <MAT>  
F;41/Active site: Arg #status predicted  
F;42/Active site: His (phosphohistidine intermediate) #status predicted  
F;92,133,167,177,191,267,322,331/Binding site: carbohydrate (Abn) (covalent) #status pre  
F;159-370,212-310,345-349/Disulfide bonds: #status predicted

Query Match 54.9%; Score 39; DB 1; Length 423;  
Best Local Similarity 70.0%; Pred. No. 29;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 PNIXTQPIPL 15  
||| |||:  
Db 132 PNISWQPIPV 141

Search completed: May 26, 2004, 18:48:02  
Job time : 1.44015 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 0.256348 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-14  
Perfect score: 71  
Sequence: 1 IALXLPNIXTQPIPL 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 43    | 60.6        | 309    | 1     | WR26_ARATH         |
| 2          | 42    | 59.2        | 426    | 1     | SYS_XANAC          |
| 3          | 42    | 59.2        | 426    | 1     | SYS_XYLFA          |
| 4          | 42    | 59.2        | 426    | 1     | SYS_XYLFT          |
| 5          | 41    | 57.7        | 1579   | 1     | SSK2_YEAST         |
| 6          | 40    | 56.3        | 258    | 1     | FLIP_ERWCA         |
| 7          | 40    | 56.3        | 373    | 1     | FUT1_RABIT         |
| 8          | 40    | 56.3        | 429    | 1     | SYS_HAEIN          |
| 9          | 39    | 54.9        | 423    | 1     | PPAL_HUMAN         |
| 10         | 39    | 54.9        | 423    | 1     | PPAL_MOUSE         |
| 11         | 39    | 54.9        | 423    | 1     | PPAL_RAT           |
| 12         | 38    | 53.5        | 416    | 1     | IL1S_RAT           |
| 13         | 38    | 53.5        | 453    | 1     | STS1_SCHPO         |
| 14         | 38    | 53.5        | 562    | 1     | LCFA_HAEIN         |
| 15         | 38    | 53.5        | 562    | 1     | LCFA_YERPE         |
| 16         | 38    | 53.5        | 651    | 1     | TOP3_HAEIN         |
| 17         | 38    | 53.5        | 766    | 1     | SFL1_YEAST         |
| 18         | 38    | 53.5        | 1087   | 1     | POM1_SCHPO         |
| 19         | 38    | 53.5        | 1239   | 1     | CHS5_USTMA         |
| 20         | 37    | 52.1        | 245    | 1     | FLIP_ECOLI         |
| 21         | 37    | 52.1        | 253    | 1     | LEP4_VIBCH         |
| 22         | 37    | 52.1        | 430    | 1     | SYS_ECOLI          |
| 23         | 37    | 52.1        | 430    | 1     | SYS_SALTY          |
| 24         | 37    | 52.1        | 561    | 1     | LCFA_ECO57         |
| 25         | 37    | 52.1        | 561    | 1     | LCFA_ECOLI         |
| 26         | 37    | 52.1        | 561    | 1     | LCFA_SALTY         |
| 27         | 37    | 52.1        | 639    | 1     | V70K_PLRVW         |
| 28         | 37    | 52.1        | 651    | 1     | PIA1_HUMAN         |
| 29         | 37    | 52.1        | 651    | 1     | PIA1_MOUSE         |
| 30         | 36    | 50.7        | 313    | 1     | LB36_ARATH         |
| 31         | 36    | 50.7        | 337    | 1     | PLSX_VIBMA         |
| 32         | 36    | 50.7        | 420    | 1     | SYS_MYCPN          |
| 33         | 36    | 50.7        | 428    | 1     | SYS_PASMU          |
|            |       |             |        |       | Q9c5t3 arabidopsis |
|            |       |             |        |       | Q8ply2 xanthomonas |
|            |       |             |        |       | Q9pb58 xylella fas |
|            |       |             |        |       | Q87bx8 xylella fas |
|            |       |             |        |       | P53599 saccharomyc |
|            |       |             |        |       | P34200 erwinia car |
|            |       |             |        |       | Q10979 oryctolagus |
|            |       |             |        |       | P43833 haemophilus |
|            |       |             |        |       | P11117 homo sapien |
|            |       |             |        |       | P24638 mus musculu |
|            |       |             |        |       | P20611 rattus norv |
|            |       |             |        |       | P43303 rattus norv |
|            |       |             |        |       | P36209 schizosacch |
|            |       |             |        |       | P46450 haemophilus |
|            |       |             |        |       | Q8zes9 yersinia pe |
|            |       |             |        |       | P43704 haemophilus |
|            |       |             |        |       | P20134 saccharomyc |
|            |       |             |        |       | Q09690 schizosacch |
|            |       |             |        |       | O13394 ustilago ma |
|            |       |             |        |       | P33133 escherichia |
|            |       |             |        |       | P27717 vibrio chol |
|            |       |             |        |       | P09156 escherichia |
|            |       |             |        |       | Q8xg98 salmonella  |
|            |       |             |        |       | P29212 escherichia |
|            |       |             |        |       | Q8xg98 salmonella  |
|            |       |             |        |       | P11622 potato leaf |
|            |       |             |        |       | O75925 homo sapien |
|            |       |             |        |       | Q88907 mus musculu |
|            |       |             |        |       | Q9fkz3 arabidopsis |
|            |       |             |        |       | Q9ra35 vibrio mari |
|            |       |             |        |       | P75107 mycoplasma  |
|            |       |             |        |       | P57836 pasteurella |

|    |      |      |      |   |            |                    |
|----|------|------|------|---|------------|--------------------|
| 34 | 36   | 50.7 | 449  | 1 | TIG_RALSO  | Q8xyp8 ralstonia s |
| 35 | 36   | 50.7 | 514  | 1 | SADI_SCHPO | Q09825 schizosacch |
| 36 | 36   | 50.7 | 601  | 1 | GLMS_PYRAE | Q8ztz0 p glucosami |
| 37 | 36   | 50.7 | 611  | 1 | ADAS_DICDI | Q06759 dictyosteli |
| 38 | 36   | 50.7 | 662  | 1 | LOX1_RABIT | P12530 oryctolagus |
| 39 | 36   | 50.7 | 662  | 1 | LOXP_RABIT | O19043 oryctolagus |
| 40 | 36   | 50.7 | 687  | 1 | YL18_CAEEL | Q11103 caenorhabdi |
| 41 | 36   | 50.7 | 809  | 1 | SYFB_RALSO | Q8xz24 ralstonia s |
| 42 | 36   | 50.7 | 1381 | 1 | Y3E7_YEAST | P34216 saccharomyc |
| 43 | 35.5 | 50.0 | 501  | 1 | DGRE_DROME | O44424 drosophila  |
| 44 | 35   | 49.3 | 193  | 1 | DCD_BIFLO  | Q8g478 bifidobacte |
| 45 | 35   | 49.3 | 331  | 1 | RDHD_HUMAN | Q8nbn7 homo sapien |

#### ALIGNMENTS

##### RESULT 1

WR26\_ARATH STANDARD; PRT; 309 AA.

ID WR26\_ARATH STANDARD; PRT; 309 AA.  
AC Q9C5T3; Q8GYK8; Q9LYQ5;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Probable WRKY transcription factor 26 (WRKY DNA-binding protein 26)  
DE (SPF1-like protein).  
GN WRKY26 OR AT5G07100 OR MOJ9.27 OR T28J14 40.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A., AND INDUCTION.  
RX MEDLINE=21342502; PubMed=11449049;  
RY Yu D., Chen C., Chen Z.;  
RT "Evidence for an important role of WRKY DNA binding proteins in the  
RT regulation of NPR1 gene expression."  
RL Plant Cell 13:1527-1540(2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=98344145; PubMed=9679202;  
RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence  
RT features of the regions of 1,381,565 bp covered by twenty one  
RT physically assigned P1 and TAC clones.";  
RL DNA Res. 5:131-145(1998).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=21016721; PubMed=11130714;  
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,  
RA Nakazaki N., Naruo K., Okumura S., Shimpo S., Takeuchi C., Wada T.,  
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,  
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,  
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Stromatt C.,  
RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,  
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,  
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,  
RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,  
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,  
RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,  
RA van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,



```
RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:823-826(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.A.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
CC responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- INDUCTION: By salicylic acid.
CC -!- SIMILARITY: Belongs to the WRKY group I family.
CC -!- SIMILARITY: Contains 2 WRKY domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF224699; AAK28309.1; -
DR EMBL; AB010697; BAB11168.1; -
DR EMBL; AL163652; CAB87266.1; -
DR EMBL; AY084692; AAM61254.1; -
DR EMBL; AK117545; BAC42206.1; -
DR PIR; T48481; T48481.
DR TRANSFAC; T05475; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS0811; WRKY; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DNA_BIND 111 176 WRKY 1.
FT DNA_BIND 228 293 WRKY 2.
FT CONFLICT 148 148 L -> P (IN REF. 5).
FT CONFLICT 165 165 I -> F (IN REF. 4).
FT CONFLICT 183 183 S -> P (IN REF. 1).
SQ SEQUENCE 309 AA; 34910 MW; B83AF11B93F3909E CRC64;

Query Match 60.6%; Score 43; DB 1; Length 309;
Best Local Similarity 57.1%; Pred. No. 1.6;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 LXLPNIXTQPIPLL 16
| | | | | | | | | |
Db 89 LFLPSMVTQPLPQL 102

RESULT 2
SYS_XANAC STANDARD; PRT; 426 AA.
AC Q8PLY2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Seryl-tRNA synthetase (EC 6.1.1.11) (Serine--tRNA ligase) (SerRS).
GN SERS OR XAC1553.
```

```
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-serine + tRNA(Ser) = AMP + diphosphate
CC + L-seryl-tRNA(Ser).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE011797; AAM36521.1; -
DR HAMAP; MF_00176; -; 1.
DR InterPro; IPR002314; tRNA-synt_2b.
DR InterPro; IPR002317; tRNA-synt_ser.
DR InterPro; IPR006195; tRNA_ligase_II.
DR Pfam; PF02403; Seryl_tRNA_N; 1.
DR Pfam; PF00587; tRNA-synt_2b; 1.
DR PRINTS; PR00981; TRNASYNTHSER.
DR TIGRfams; TIGR00414; serS; 1.
DR PROSITE; PS50862; AA tRNA_LIGASE_II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 426 AA; 47118 MW; B72EBA3070F6F5A4 CRC64;

Query Match 59.2%; Score 42; DB 1; Length 426;
Best Local Similarity 53.3%; Pred. No. 3.6;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 15
| | | | | | | | | |
Db 100 IALGLPNLPAEGVPL 114

RESULT 3
SYS_XYLEA STANDARD; PRT; 426 AA.
AC Q9PB58;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Seryl-tRNA synthetase (EC 6.1.1.11) (Serine--tRNA ligase) (SerRS).
GN SERS OR XF2286.
OS Xylella fastidiosa.
```



AC P53599;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE MAP kinase kinase SSK2 (EC 2.7.1.-) (Suppressor of sensor  
 DE kinase 2).  
 GN SSK2 OR YNR031C OR N3276.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95350642; PubMed=7624781;  
 RA Maeda T., Takekawa M., Saito H.;  
 RT "Activation of yeast PBS2 MAPKK by MAPKKs or by binding of an SH3-  
 RT containing osmosensor.";  
 RL Science 269:554-558(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Pohl T.M.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Kinase involved in a signal transduction pathway that is  
 CC activated by changes in the osmolarity of the extracellular  
 CC environment. Activates the PBS2 MAP kinase by  
 CC phosphorylation.  
 CC -!- SUBUNIT: Interacts with by SSK1.  
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. MAP  
 CC kinase kinase subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L41927; AAC41665.1; -.  
 CC EMBL; 271646; CAA96311.1; -.  
 CC PIR; S59801; S59801.  
 CC GerMOnline; 143376; -.  
 CC SGD; S0005314; SSK2.  
 CC GO; GO:0005829; C:cytosol; IDA.  
 CC GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; IDA.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR008271; Ser\_thr\_kinase.  
 CC InterPro; IPR002290; Ser\_thr\_kinase.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 1.  
 CC SMART; SM00220; S\_TKc; 1.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 CC Transferrase; Serine/threonine-protein kinase; ATP-binding.  
 KW DOMAIN 1266 1558 PROTEIN\_KINASE.  
 FT NP\_BIND 1272 1280 ATP (BY SIMILARITY).  
 FT BINDING 1295 1295 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1390 1390 BY SIMILARITY.  
 SQ SEQUENCE 1579 AA; 180526 MW; 91A07A195CB3772E CRC64;  
 Query Match 57.7%; Score 41; DB 1; Length 1579;  
 Best Local Similarity 54.5%; Pred. No. 24;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 LPNIXTQPIPL 15  
 Db 132 LPKLSTQPVVPV 142  
 RESULT 6  
 FLIP\_ERWCA STANDARD; PRT; 258 AA.  
 ID FLIP\_ERWCA

AC P34200;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Flagellar biosynthetic protein flip (Flagellar biosynthetic protein  
 DE mopC).  
 GN FLIP OR MOPC.  
 OS Erwinia carotovora.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pectobacterium.  
 OX NCBI\_TaxID=554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=SCRI193;  
 RA MEDLINE=94018629; PubMed=8412685;  
 RA Mulholland V., Hinton J.C.D., Sidebotham J., Toth I.K.,  
 RA Hyman L.J., Perombelon M.C.M., Reeves P.J., Salmond G.P.C.;  
 RT "A pleiotropic reduced virulence (Rvi-) mutant of Erwinia carotovora  
 RT subspecies atroseptica is defective in flagella assembly proteins  
 RT that are conserved in plant and animal bacterial pathogens.";  
 RL Mol. Microbiol. 9:343-356(1993).  
 CC -!- FUNCTION: Plays a role in the flagellum-specific transport system  
 CC (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE FLIP/MOPC/SPAP FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X72969; CAA51476.1; -.  
 CC PIR; S35276; S35276.  
 CC InterPro; IPR005837; Flip.  
 CC InterPro; IPR005838; TypeIII\_P.  
 CC Pfam; PF00813; Flip; 1.  
 CC PRINTS; PR01302; TYPE3IMPPTOT.  
 CC ProDom; PD002586; TypeIII\_P; 1.  
 CC TIGRfams; TIGR01103; Flip; 1.  
 CC PROSITE; PS01060; FLIP\_1; 1.  
 CC PROSITE; PS01061; FLIP\_2; 1.  
 KW Flagellum; Transmembrane.  
 FT TRANSMEM 20 40 POTENTIAL.  
 FT TRANSMEM 64 87 POTENTIAL.  
 FT TRANSMEM 98 118 POTENTIAL.  
 FT TRANSMEM 194 213 POTENTIAL.  
 FT TRANSMEM 220 242 POTENTIAL.  
 SQ SEQUENCE 258 AA; 28557 MW; BBA2BE8389900A5C8 CRC64;  
 Query Match 56.3%; Score 40; DB 1; Length 258;  
 Best Local Similarity 70.0%; Pred. No. 4.8;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 LPNIXTQPIP 14  
 Db 36 LPGIVTQPLP 45  
 RESULT 7  
 FUT1\_RABIT STANDARD; PRT; 373 AA.  
 ID FUT1\_RABIT  
 AC Q10979;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Galactoside 2-alpha-L-fucosyltransferase 1 (EC 2.4.1.69) (GDP-L-  
 DE fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase 1).  
 DE (Alpha(1,2)Ft 1) (Fucosyltransferase 1).  
 GN FUT1 OR RFT-I.  
 OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;

RX MEDLINE=95238380; PubMed=7721792;  
RA Hitoshi S., Kusunoki S., Kanazawa I., Tsuji S.;

RT "Molecular cloning and expression of two types of rabbit beta-  
RT galactoside alpha 1,2-fucosyltransferase.";  
RL J. Biol. Chem. 270:8844-8850(1995).

CC -|- CATALYTIC ACTIVITY: GDP-L-fucose + beta-D-galactosyl-R = GDP +  
CC alpha-L-fucosyl-1,2-beta-D-galactosyl-R.  
CC -|- PATHWAY: Glycosylation.

CC -|- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC form in trans cisternae of Golgi.

CC -|- TISSUE SPECIFICITY: Brain.

CC -|- MISCELLANEOUS: THERE ARE TWO GENES (FUT1 AND FUT2) WHICH ENCODE  
CC GALACTOSIDE 2-L-FUCOSYLTRANSFERASE. THEY ARE EXPRESSED IN A

CC TISSUE-SPECIFIC MANNER WITH EXPRESSION RESTRICTED TO CELLS OF  
CC MESODERMAL OR ENDODERMAL ORIGIN RESPECTIVELY.

CC -|- SIMILARITY: Belongs to the glycosyltransferase family 11.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X80226; CAA56513.1; -.  
DR PIR; A56392; A56392.

DR InterPro; IPR002516; Glyco\_trans\_11.  
DR Pfam; PF01531; Glyco\_transf\_11; 1.

KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;  
KW Signal-anchor; Golgi stack.

FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 13 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).

FT DOMAIN 30 373 LUMENAL, CATALYTIC (POTENTIAL).  
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 373 AA; 42098 MW; 0A47A1786231525C CRC64;

Query Match 56.3%; Score 40; DB 1; Length 373;  
Best Local Similarity 50.0%; Pred. No. 7.3;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIP 14  
Db 37 LALSPLCLERQVPV 50

RESULT 8

SYS\_HAEIN

ID SYS\_HAEIN STANDARD; PRT; 429 AA.

AC P43833;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Seryl-trNA synthetase (EC 6.1.1.11) (Serine--trNA ligase) (SerRS).

GN SERS OR HI0110.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI\_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Rd / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd.";

RL Science 269:496-512(1995).

CC -|- CATALYTIC ACTIVITY: ATP + L-serine + trNA(Ser) = AMP + diphosphate  
CC + L-seryl-trNA(Ser).

CC -|- SUBUNIT: Homodimer (By similarity).

CC -|- SUBCELLULAR LOCATION: Cytoplasmic.

CC -|- SIMILARITY: Belongs to class-II aminoacyl-trNA synthetase family.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U32696; AAC21787.1; -.

DR PIR; I64048; I64048.

DR HSSP; P34945; 1SER.

DR TIGR; HI0110; -.

DR HAMAP; MF\_00176; -; 1.

DR InterPro; IPR002314; trNA-synt\_2b.

DR InterPro; IPR002317; trNA-synt\_ser.

DR InterPro; IPR006195; trNA\_ligase\_II.

DR Pfam; PF02403; Seryl\_trNA\_N; 1.

DR Pfam; PF00587; trNA-synt\_2b; 1.

DR PRINTS; PR00981; TRNASYNTHSER.

DR TIGRFAMs; TIGR00414; sers; 1.

DR PROSITE; PS50862; AA TRNA LIGASE II; 1.

KW Aminoacyl-trNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW Complete proteome.

SQ SEQUENCE 429 AA; 47721 MW; 763967DFCD9838ED CRC64;

Query Match 56.3%; Score 40; DB 1; Length 429;  
Best Local Similarity 46.7%; Pred. No. 8.5;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIP 15

Db 100 IALSIPNLPADDEVPL 114

RESULT 9

PPAL\_HUMAN

ID PPAL\_HUMAN STANDARD; PRT; 423 AA.

AC P11117; Q9BTU7;

DT 01-JUL-1989 (Rel. 11, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).

GN ACP2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Placenta;

RX MEDLINE=89052645; PubMed=3191910;

RA Pohlmann R., Krentler C., Schmidt B., Schroeder W., Lorkowski G.,

RA Culley J., Mersmann G., Geier C., Waheed A., Gottschalk S.,

RA Grzeschik K.H., Hasikik A., von Figura K.;

RT "Human lysosomal acid phosphatase: cloning, expression and

RT chromosomal assignment.";



RL EMBO J. 7:2343-2350(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Leukocyte;  
RA MEDLINE=89377828; PubMed=2776754;  
RX Geier C., von Figura K., Pohlmann R.;  
RT "Structure of the human lysosomal acid phosphatase gene.";  
RL Eur. J. Biochem. 183:611-616(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
CC alcohol + phosphate.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X12548; CAA31064.1; -  
DR EMBL; X15525; CAA33542.1; -  
DR EMBL; X15526; CAA33542.1; JOINED.  
DR EMBL; X15527; CAA33542.1; JOINED.  
DR EMBL; X15528; CAA33542.1; JOINED.  
DR EMBL; X15529; CAA33542.1; JOINED.  
DR EMBL; X15530; CAA33542.1; JOINED.  
DR EMBL; X15531; CAA33542.1; JOINED.  
DR EMBL; X15532; CAA33542.1; JOINED.  
DR EMBL; X15533; CAA33542.1; JOINED.  
DR EMBL; X15534; CAA33542.1; JOINED.  
DR EMBL; X15535; CAA33542.1; JOINED.  
DR EMBL; BC003160; AAB03160.1; -  
DR PIR; S06167; S06167.  
DR HSSP; P20646; IRPA.  
DR Genew; HGNC:123; ACP2.  
DR MIM; 171650; -  
DR GO; GO:0016021; C:integral to membrane; TAS.  
DR GO; GO:0003993; F:acid phosphatase activity; TAS.  
DR InterPro; IPR000560; HisAc\_phsphtse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; signal; Glycoprotein; Lysosome.  
FT SIGNAL 1 30  
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
FT DISULFID 159 370 BY SIMILARITY.  
FT DISULFID 212 310 BY SIMILARITY.  
FT -----

FT DISULFID 345 349 BY SIMILARITY.  
FT ACT\_SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT ACT\_SITE 84 84 BY SIMILARITY.  
FT ACT\_SITE 286 286 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 29 29 Q -> R (IN REF. 1 AND 2).  
SQ SEQUENCE 423 AA; 48316 MW; 1A2CBFB40ED3724B CRC64;  
  
Query Match 54.9%; Score 39; DB 1; Length 423;  
Best Local Similarity 70.0%; Pred. No. 13;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 6 PNIXTQPIPL 15  
Db 132 PNISWQPIPV 141  
|||||  
  
RESULT 10  
PPAL MOUSE  
ID PPAL MOUSE STANDARD; PRT; 423 AA.  
AC P24638; Q8QZT5;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
GN ACP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE OF 3-423 FROM N.A.  
RX MEDLINE=91282986; PubMed=2059337;  
RA Geier C., von Figura K., Pohlmann R.;  
RT "Molecular cloning of the mouse lysosomal acid phosphatase.";  
RL Biol. Chem. Hoppe-Seyler 372:301-304(1991).  
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
CC alcohol + phosphate.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration







Db 92 LTLPGVTRQGLPL 104

RESULT 14

```

LCFA_HAEIN          STANDARD;          PRT;          562 AA.
ID LCFA_HAEIN
AC P46450;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) (Long-chain acyl-CoA
DE synthetase).
GN FADD OR HI0390.1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION.
RA Koonin E.V., Rudd K.E.;
RL Submitted (SEP-1995) to Swiss-Prot.
CC -!- FUNCTION: Esterification, concomitant with transport, of exogenous
CC long-chain fatty acids into metabolically active CoA thioesters
CC for subsequent degradation or incorporation into phospholipids (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + a long-chain carboxylic acid + CoA = AMP
CC + diphosphate + an acyl-CoA.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SUBUNIT: Homodimer (Probable).
CC -!- SUBCELLULAR LOCATION: Partially membrane-associated (Potential).
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.

```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

```

EMBL; U32722; AAC22049.1; -
HSP; P08659; 1LCI.
TIGR; HI0390.1; -
InterPro; IPR000873; AMP-bind.
Pfam; PF00501; AMP-binding; 1.
PROSITE; PS00455; AMP_BINDING; 1.
Ligase; Fatty acid metabolism; Magnesium; Membrane; Complete proteome.
SQ SEQUENCE 562 AA; 63478 MW; 74BE7C8E9D711F12 CRC64;

```

Query Match 53.5%; Score 38; DB 1; Length 562;  
Best Local Similarity 46.7%; Pred. No. 27;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 15

Db 77 VALMMPNLLQYPAL 91

RESULT 15

```

LCFA_YERPE          STANDARD;          PRT;          562 AA.
ID LCFA_YERPE
AC Q8ZES9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) (Long-chain acyl-CoA
DE synthetase).
GN FADD OR YPO2074 OR Y2236.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- FUNCTION: Esterification, concomitant with transport, of exogenous
CC long-chain fatty acids into metabolically active CoA thioesters
CC for subsequent degradation or incorporation into phospholipids (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + a long-chain carboxylic acid + CoA = AMP
CC + diphosphate + an acyl-CoA.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SUBUNIT: Homodimer (Probable).
CC -!- SUBCELLULAR LOCATION: Partially membrane-associated (Potential).
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.

```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

```

EMBL; AJ414151; CAC90886.1; -
EMBL; AE013826; AAM85796.1; ALT_INIT.
PIR; AB0253; AB0253.
InterPro; IPR000873; AMP-bind.
Pfam; PF00501; AMP-binding; 1.
PROSITE; PS00455; AMP_BINDING; 1.
Ligase; Fatty acid metabolism; Magnesium; ATP-binding; Membrane;
Complete proteome.
FT NP_BIND 213 224 ATP (PROBABLE).
FT VARIANT 504 504 N -> K (IN STRAIN KIM5).
SQ SEQUENCE 562 AA; 62639 MW; 812A872B3713D3FD CRC64;

```

Query Match 53.5%; Score 38; DB 1; Length 562;



Best Local Similarity 46.7%; Pred. No. 27;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTOPIPL 15  
:|:|:|:|:|:|  
Db 77 VALMPNLLQYPIAL 91

Search completed: May 26, 2004, 18:41:22  
Job time : 2.25635 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 1.2769 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-14  
Perfect score: 71  
Sequence: 1 IALXLPNIXTQPIPLL 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 65    | 91.5        | 920    | 11 Q8CB72 | Q8cb72 mus musculu |
| 2          | 65    | 91.5        | 961    | 11 Q80YQ6 | Q80yq6 mus musculu |
| 3          | 65    | 91.5        | 968    | 11 O88622 | O88622 mus musculu |
| 4          | 65    | 91.5        | 972    | 11 Q9QYM2 | Q9qym2 rattus norv |
| 5          | 65    | 91.5        | 976    | 4 Q9Y4W7  | Q9y4w7 homo sapien |
| 6          | 65    | 91.5        | 976    | 4 Q86W56  | Q86w56 homo sapien |
| 7          | 65    | 91.5        | 976    | 4 Q7Z742  | Q7z742 homo sapien |
| 8          | 65    | 91.5        | 977    | 6 Q02776  | Q02776 bos taurus  |
| 9          | 46    | 64.8        | 723    | 5 Q960N8  | Q960n8 drosophila  |
| 10         | 46    | 64.8        | 768    | 5 O46043  | O46043 drosophila  |
| 11         | 44    | 62.0        | 317    | 5 O16510  | O16510 caenorhabdi |
| 12         | 43    | 60.6        | 524    | 5 O16506  | O16506 caenorhabdi |
| 13         | 43    | 60.6        | 1256   | 16 Q8YVT6 | Q8yvt6 anabaena sp |
| 14         | 42    | 59.2        | 203    | 16 Q8AAE5 | Q8aae5 bacteroides |
| 15         | 42    | 59.2        | 249    | 5 Q8T306  | Q8t306 tribolium c |
| 16         | 42    | 59.2        | 513    | 12 Q83869 | Q83869 narcissus l |

|    |      |      |      |           |                     |
|----|------|------|------|-----------|---------------------|
| 17 | 41.5 | 58.5 | 567  | 16 Q8A6R3 | Q8a6r3 bacteroides  |
| 18 | 41   | 57.7 | 124  | 5 Q9XW97  | Q9xw97 caenorhabdi  |
| 19 | 41   | 57.7 | 432  | 16 Q882S3 | Q882s3 pseudomonas  |
| 20 | 41   | 57.7 | 499  | 16 Q82M67 | Q82m67 streptomyce  |
| 21 | 41   | 57.7 | 1072 | 3 Q94537  | Q94537 schizosacch  |
| 22 | 41   | 57.7 | 2096 | 16 Q7TWN7 | Q7tnw7 mycobacteri  |
| 23 | 41   | 57.7 | 3157 | 16 O50378 | O50378 mycobacteri  |
| 24 | 40   | 56.3 | 136  | 5 Q9NA23  | Q9na23 caenorhabdi  |
| 25 | 40   | 56.3 | 336  | 10 Q7XPK3 | Q7xpk3 oryza sativ  |
| 26 | 40   | 56.3 | 368  | 10 Q8S164 | Q8s164 oryza sativ  |
| 27 | 40   | 56.3 | 551  | 11 Q8VCA0 | Q8vca0 mus musculu  |
| 28 | 40   | 56.3 | 654  | 12 Q9DSF7 | Q9dsp7 saint croix  |
| 29 | 40   | 56.3 | 774  | 5 Q9VU73  | Q9vj73 drosophila   |
| 30 | 39   | 54.9 | 259  | 16 Q988L9 | Q988l9 rhizobium l  |
| 31 | 39   | 54.9 | 296  | 2 Q9ADW8  | Q9adw8 ehrlichia c  |
| 32 | 39   | 54.9 | 305  | 16 Q89V11 | Q89v11 bradyrhizob  |
| 33 | 39   | 54.9 | 508  | 16 Q7UHI2 | Q7uh12 rhodopirell  |
| 34 | 39   | 54.9 | 526  | 10 Q9LZ26 | Q9l226 arabidopsis  |
| 35 | 39   | 54.9 | 526  | 10 Q8GWI9 | Q8gw19 arabidopsis  |
| 36 | 39   | 54.9 | 585  | 16 Q8PQAL | Q8ppqal xanthomonas |
| 37 | 39   | 54.9 | 893  | 10 Q9SXH2 | Q9sxh2 ipomoea bat  |
| 38 | 39   | 54.9 | 943  | 3 Q8X0B4  | Q8x0b4 neurospora   |
| 39 | 39   | 54.9 | 997  | 16 Q8D4H3 | Q8d4h3 vibrio vuln  |
| 40 | 39   | 54.9 | 1031 | 2 Q9AEQ8  | Q9aeq8 vibrio vuln  |
| 41 | 39   | 54.9 | 1031 | 2 Q9AHK2  | Q9ahk2 vibrio vuln  |
| 42 | 39   | 54.9 | 1800 | 5 Q8IEI9  | Q8iei9 plasmodium   |
| 43 | 38   | 53.5 | 36   | 6 Q9GKJ9  | Q9gkj9 sus scrofa   |
| 44 | 38   | 53.5 | 138  | 1 Q9UXE2  | Q9uxe2 sulfolobus   |
| 45 | 38   | 53.5 | 252  | 4 Q8N420  | Q8n420 homo sapien  |

ALIGNMENTS

RESULT 1

Q8CB72 ID Q8CB72 PRELIMINARY; PRT; 920 AA.  
AC Q8CB72;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Poly.  
GN PARC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; Par9.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 91.5%; Score 65; DB 11; Length 920;  
Best Local Similarity 87.5%; Pred.No. 0.0017;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
Db 593 IALCLPNICTQPIPLL 608

RESULT 2

Q80YQ6

ID Q80YQ6 PRELIMINARY; PRT; 961 AA.  
AC Q80YQ6;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050892; AAH50892.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

Query Match 91.5%; Score 65; DB 11; Length 961;  
Best Local Similarity 87.5%; Pred. No. 0.0018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
||| ||||| ||||| ||  
Db 593 IALCLPNICTQPIPLL 608

RESULT 3  
O88622 ID Q88622 PRELIMINARY; PRT; 968 AA.  
AC O88622;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93380098; PubMed=10449915;  
RA Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
RT hybridization";  
RL Cytogenet. Cell Genet. 85:269-270(1999).  
DR EMBL; AF079557; AAC28735.1; -.  
DR MGD; MGI:1347094; Parg.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;

Query Match 91.5%; Score 65; DB 11; Length 968;  
Best Local Similarity 87.5%; Pred. No. 0.0018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
||| ||||| ||||| ||  
Db 592 IALCLPNICTQPIPLL 607

RESULT 4  
Q9QYM2 ID Q9QYM2 PRELIMINARY; PRT; 972 AA.  
AC Q9QYM2;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
GN PARG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BUF; TISSUE=Colon;  
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,  
RA Aoki Y., Nakgama H., Sugimura T.;  
RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase  
RT (Parg).";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB019366; BAA87901.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;

Query Match 91.5%; Score 65; DB 11; Length 972;  
Best Local Similarity 87.5%; Pred. No. 0.0018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
||| ||||| ||||| ||  
Db 596 IALCLPNICTQPIPLL 611

RESULT 5  
Q9Y4W7 ID Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
AC Q9Y4W7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93380098; PubMed=10449915;  
RA Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
RT hybridization";  
RL Cytogenet. Cell Genet. 85:269-270(1999).  
DR EMBL; AF005043; AAB61614.1; -.  
DR Genew; HGNC:8605; PARG.  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;

Query Match 91.5%; Score 65; DB 4; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.0019;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
||| ||||| ||||| ||  
Db 600 IALCLPNICTQPIPLL 615

RESULT 6

Q86W56  
ID Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
  
Query Match 91.5%; Score 65; DB 4; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.0019;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 IALXLPNIXTQPIPLL 16  
Db 600 IALCLPNICTQPIPLL 615  
  
RESULT 7  
Q7Z742 PRELIMINARY; PRT; 976 AA.  
ID Q7Z742  
AC Q7Z742;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC052966; AAH52966.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;  
  
Query Match 91.5%; Score 65; DB 4; Length 976;  
Best Local Similarity 87.5%; Pred. No. 0.0019;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 IALXLPNIXTQPIPLL 16  
Db 600 IALCLPNICTQPIPLL 615  
  
RESULT 8  
O02776 PRELIMINARY; PRT; 977 AA.  
ID O02776  
AC O02776;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN BPARG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97277328; PubMed=9115250;  
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-  
RT ribose) glycohydrolase.";  
RL J. Biol. Chem. 272:11895-11901 (1997).  
DR EMBL; U78975; AAB53370.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;  
  
Query Match 91.5%; Score 65; DB 6; Length 977;  
Best Local Similarity 87.5%; Pred. No. 0.0019;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 IALXLPNIXTQPIPLL 16  
Db 601 IALCLPNICTQPIPLL 616  
  
RESULT 9  
Q960N8 PRELIMINARY; PRT; 723 AA.  
ID Q960N8  
AC Q960N8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE LD42380p.  
GN PARG OR EG:114E2.1 OR CG2864.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.



```
DR EMBL; AY051955; AAK93379.1; -.
DR FlyBase; FBgn0023216; Parg.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;

Query Match 64.8%; Score 46; DB 5; Length 723;
Best Local Similarity 50.0%; Pred. No. 5.3;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16
Db 174 LALRLPDLIQSPVPLL 189

RESULT 10
O46043 PRELIMINARY; PRT; 768 AA.
AC O46043;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PARG protein.
GN PARG OR EG:114E2.1 OR CG2864.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
```

```
RA Ame J.-C., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding Drosophila
RT poly(ADP-ribose) glycohydrolase."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 46-768 FROM N.A.
RA Murphy L., Harris D., Barrell B.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 46-768 FROM N.A.
RA Benos P.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003428; AAF45886.1; -.
DR EMBL; AF079556; AAC28734.1; -.
DR EMBL; Z98254; CAB10913.1; -.
DR FlyBase; FBgn0023216; Parg.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;

Query Match 64.8%; Score 46; DB 5; Length 768;
Best Local Similarity 50.0%; Pred. No. 5.7;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16
Db 219 LALRLPDLIQSPVPLL 234

RESULT 11
O16510 PRELIMINARY; PRT; 317 AA.
AC O16510;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE C03A7.13 protein.
GN C03A7.13.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans."
RN Nature 368:32-38 (1994).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Greco T., Bradshaw H., Elliott G.;
RT "The sequence of C. elegans cosmid C03A7."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN EMBL; AF016451; AAB65994.1; -.
RP
```

DR PIR; T31897; T31897.  
DR WormPep; C03A7.13; CE07871.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR InterPro; IPR002048; EF-hand.  
DR PROSITE; PS00018; EF\_HAND; 1.  
SQ SEQUENCE 317 AA; 35452 MW; 8F9CB24E185DE488 CRC64;  
  
Query Match 62.0%; Score 44; DB 5; Length 317;  
Best Local Similarity 46.7%; Pred. No. 5.2;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 IALXLPNIXTQPIPL 15  
Db 173 VTGMLPNIASQVPL 187  
  
RESULT 12  
O16506 PRELIMINARY; PRT; 524 AA.  
AC O16506;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein C03A7.11.  
GN C03A7.11.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RT Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX Greco T., Bradshaw H., Elliott G.;  
RT "The sequence of C. elegans cosmid C03A7.";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF016451; AAB66005.2; -.  
DR PIR; T31893; T31893.  
DR WormPep; C03A7.11; CE32564.  
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . .; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR002213; UDP\_gluco\_trans.  
DR Pfam; PF00201; UDPGT; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 524 AA; 59372 MW; C69D9417434EF130 CRC64;  
  
Query Match 60.6%; Score 43; DB 5; Length 524;  
Best Local Similarity 46.7%; Pred. No. 14;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 IALXLPNIXTQPIPL 15  
Db 173 VTGMLPNIASQVPL 187  
  
RESULT 13  
Q8YVT6  
ID Q8YVT6 PRELIMINARY; PRT; 1256 AA.  
AC Q8YVT6;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein All1888.  
GN All1888.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003587; BAB73587.1; -.  
DR PIR; AB2042; AB2042.  
DR InterPro; IPR008938; ARM.  
DR InterPro; IPR007111; NACHT\_NTPase.  
DR Pfam; PF05729; NACHT; 1.  
DR PROSITE; PS50837; NACHT; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1256 AA; 147253 MW; DDFB5E19241BAC6 CRC64;  
  
Query Match 60.6%; Score 43; DB 16; Length 1256;  
Best Local Similarity 66.7%; Pred. No. 36;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 5 LPNIXTQPIPL 16  
Db 333 LDNVISQPIPL 344  
  
RESULT 14  
Q8AAE5  
ID Q8AAE5 PRELIMINARY; PRT; 203 AA.  
AC Q8AAE5;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Transcriptional regulator.  
GN BT0520.  
OS Bacteroides thetaiotaomicron.  
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
OC Bacteroidaceae; Bacteroides.  
OX NCBI\_TaxID=818;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VPI-5482 / ATCC 29148;  
RX MEDLINE=22550858; PubMed=12663928;  
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,  
RA Chiang H.C., Hooper L.V., Gordon J.I.;  
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";  
RL Science 299:2074-2076(2003).  
DR EMBL; AE016928; AAO75627.1; -.  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004672; F:protein kinase activity; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0000156; F:two-component response regulator activity; IEA.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.  
DR InterPro; IPR000792; HTH\_LuxR.  
DR InterPro; IPR000719; Prot\_kinase.  
DR Pfam; PF00196; GerE; 1.  
DR PRINTS; PR00038; HTHLUXR.  
DR ProDom; PD000307; HTH\_LuxR; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR ProDom; PD000039; Response reg; 1.  
DR SMART; SM00421; HTH\_LUXR; 1.

```
DR PROSITE; PS00622; HTH_LUXR_FAMILY; 1.
KW Complete proteome.
SQ SEQUENCE 203 AA; 22763 MW; 4DC2F80D18D4F510 CRC64;

Query Match 59.2%; Score 42; DB 16; Length 203;
Best Local Similarity 66.7%; Pred. No. 7.7;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPLL 16
   |||: ||| ||
Db 29 LPNVKVQPIELL 40

RESULT 15
Q8T306 PRELIMINARY; PRT; 249 AA.
AC Q8T306;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Helix-loop-helix transcription factor.
GN HAIRY.
OS Tribolium castaneum (Red flour beetle).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
OC Tenebrionidae; Tribolium.
OX NCBI_TaxID=7070;
RN [1]
RP SEQUENCE FROM N.A.
RA Eckert C., Wolff C., Wimmer E., Tautz D.;
RT "Functional analysis of the regulatory region of the pair-rule gene
RT hairy in Tribolium suggests regulatory divergence inspite of conserved
RT expression.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
DR EMBL; AJ457831; CAD29886.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR003650; Orange.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00511; ORANGE; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50888; HLH_2; 1.
SQ SEQUENCE 249 AA; 27252 MW; 364BB299EEBF429B CRC64;

Query Match 59.2%; Score 42; DB 5; Length 249;
Best Local Similarity 56.2%; Pred. No. 9.6;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16
   ||| ||| :|||
Db 184 IALVLPQTQASPLPLL 199
```

Search completed: May 26, 2004, 18:46:14  
Job time : 3.2769 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 4.80653 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977  
Perfect score: 1597  
Sequence: 1 KTLFCYFRRVTEKPTGLVT.....YHAVESCAETADHSGQRTGT 300

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description         |
|------------|-------|-------------|--------|--------------|---------------------|
| 1          | 91.5  | 5.7         | 261    | 1 LAS1_HUMAN | Q14847 homo sapien  |
| 2          | 91.5  | 5.7         | 491    | 1 K2M2_SHEEP | P15241 ovis aries   |
| 3          | 90.5  | 5.7         | 263    | 1 LAS1_MOUSE | Q61792 mus musculus |
| 4          | 89    | 5.6         | 727    | 1 IF2M_MOUSE | Q91yJ5 mus musculus |
| 5          | 86    | 5.4         | 153    | 1 ENPP_BBP21 | P27358 bacterioph   |
| 6          | 86    | 5.4         | 153    | 1 ENPP_LAMB2 | P00726 bacterioph   |
| 7          | 85    | 5.3         | 523    | 1 RPB2_HALN1 | P15352 halobacteri  |
| 8          | 84.5  | 5.3         | 304    | 1 CYSD_XYLFA | Q9pd79 xyella fas   |
| 9          | 84    | 5.3         | 2298   | 1 YCF2_LOTJA | Q9b1k6 lotus japon  |
| 10         | 83.5  | 5.2         | 153    | 1 RZPD_ECOLI | P75719 escherichia  |
| 11         | 83    | 5.2         | 1135   | 1 RBL2_RAT   | O55081 rattus norv  |
| 12         | 82.5  | 5.2         | 624    | 1 HTPG_CLOAB | Q97e05 clostridium  |
| 13         | 82.5  | 5.2         | 854    | 1 AKA3_BOVIN | O77797 bos taurus   |
| 14         | 82.5  | 5.2         | 998    | 1 BGAL_LACLA | Q48727 lactococcus  |
| 15         | 81.5  | 5.1         | 982    | 1 CBLB_HUMAN | Q13191 homo sapien  |
| 16         | 81    | 5.1         | 1575   | 1 IQG2_HUMAN | Q13576 homo sapien  |
| 17         | 79.5  | 5.0         | 618    | 1 IORA_METTH | O27880 methanobact  |
| 18         | 79    | 4.9         | 721    | 1 HS90_THEPA | P24724 theileria p  |
| 19         | 78.5  | 4.9         | 317    | 1 YCQ5_YEAST | P25616 saccharomyc  |
| 20         | 78.5  | 4.9         | 514    | 1 TSSP_HUMAN | Q9nqe7 homo sapien  |
| 21         | 78.5  | 4.9         | 761    | 1 EFG1_YEAST | P25039 saccharomyc  |
| 22         | 78.5  | 4.9         | 989    | 1 IDE_DROME  | P22817 drosophila   |
| 23         | 78    | 4.9         | 405    | 1 SECY_ANTSP | Q37143 antithamnio  |
| 24         | 78    | 4.9         | 469    | 1 ORB6_SCHPO | O13310 schizosacch  |
| 25         | 78    | 4.9         | 703    | 1 HS83_TRYBB | P12861 trypanosoma  |
| 26         | 77.5  | 4.9         | 252    | 1 HIS6_RHOCA | O30724 rhodobacter  |
| 27         | 77.5  | 4.9         | 283    | 1 YE34_DEIRA | Q9ruf2 deinococcus  |
| 28         | 77.5  | 4.9         | 314    | 1 MAG3_HUMAN | P43357 homo sapien  |
| 29         | 77.5  | 4.9         | 384    | 1 RECF_BRUME | Q8yed7 brucella me  |
| 30         | 77.5  | 4.9         | 404    | 1 ISPG_TREPA | O83460 treponema p  |
| 31         | 77.5  | 4.9         | 569    | 1 YXID_BACSU | P42296 bacillus su  |
| 32         | 77    | 4.8         | 344    | 1 FLIM_TREPA | P74927 treponema p  |
| 33         | 77    | 4.8         | 414    | 1 YF24_MYCTU | Q50583 mycobacteri  |

|    |      |     |      |              |                    |
|----|------|-----|------|--------------|--------------------|
| 34 | 76.5 | 4.8 | 314  | 1 MAG6_HUMAN | P43360 homo sapien |
| 35 | 76.5 | 4.8 | 343  | 1 LPSD_RHIME | Q9r9n0 rhizobium m |
| 36 | 76.5 | 4.8 | 477  | 1 NIFD_METMP | P71526 methanococc |
| 37 | 76.5 | 4.8 | 520  | 1 UBP3_MOUSE | Q91w36 mus musculu |
| 38 | 76.5 | 4.8 | 649  | 1 RA32_SCHPO | Q09683 schizosacch |
| 39 | 76   | 4.8 | 366  | 1 HI82_BACCR | Q81c43 bacillus ce |
| 40 | 76   | 4.8 | 593  | 1 CDRI_SCHPO | P07334 schizosacch |
| 41 | 76   | 4.8 | 627  | 1 PLSL_HUMAN | P13796 homo sapien |
| 42 | 76   | 4.8 | 657  | 1 UVRB_CAMJE | Q9ppm7 campylobact |
| 43 | 76   | 4.8 | 837  | 1 HFC3_HAEIN | P45998 haemophilus |
| 44 | 76   | 4.8 | 839  | 1 OS94_HUMAN | O95757 homo sapien |
| 45 | 76   | 4.8 | 4447 | 1 PKSK_BACSU | P40803 bacillus su |

ALIGNMENTS

RESULT 1  
LAS1\_HUMAN  
ID LAS1\_HUMAN STANDARD; PRT; 261 AA.  
AC Q14847; Q96ED2;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE LIM and SH3 domain protein 1 (LASP-1) (MLN 50).  
GN LASP1 OR MLN50.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=96039245; PubMed=7490069;  
RA Tomasetto C., Regnier C., Moog-Lutz C., Mattei M.-G., Chenard M.-P.,  
RA Lidereau R., Basset P., Rio M.-C.;  
RT "Identification of four novel human genes amplified and overexpressed  
RT in breast carcinoma and localized to the q11-q21.3 region of  
RT chromosome 17.";  
RL Genomics 28:367-376(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Euetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshikiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP DOMAINS.  
RX MEDLINE=96033982; PubMed=7589475;  
RA Tomasetto C., Moog-Lutz C., Regnier C.H., Schreiber V., Basset P.,  
RA Rio M.-C.;  
RT "Lasp-1 (MLN 50) defines a new LIM protein subfamily characterized by  
RT the association of LIM and SH3 domains.";  
RL FEBS Lett. 373:245-249(1995).  
CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.



```

CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chromancer/Genes/Lasp1ID203.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X82456; CAA57833.1; -.
CC EMBL; BC012460; AAH12460.1; -.
CC HSSP; P80171; 1ZFO.
CC SWISS-2DPAGE; Q14847; HUMAN.
CC Genew; HGNC:6513; LASP1.
CC MIM; 602920; -.
CC GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.
CC InterPro; IPR001781; LIM.
CC InterPro; IPR000900; Nebulin.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00412; LIM; 1.
CC Pfam; PF00880; Nebulin; 2.
CC Pfam; PF00018; SH3; 1.
CC PRINTS; PR00452; SH3DOMAIN.
CC ProDom; PD000094; LIM; 1.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00132; LIM; 1.
CC SMART; SM00227; NEBU; 2.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS00478; LIM_DOMAIN_1; 1.
CC PROSITE; PS50023; LIM_DOMAIN_2; 1.
CC PROSITE; PS50002; SH3; 1.
CC LIM domain; Metal-binding; Zinc; SH3 domain.
FT DOMAIN 5 56 LIM.
FT DOMAIN 202 261 SH3.
FT DOMAIN 201 204 POLY-GLY.
FT CONFLICT 79 79 E -> R (IN REF. 2).
FT CONFLICT 210 210 V -> A (IN REF. 2).
FT CONFLICT 220 220 E -> A (IN REF. 2).
SQ SEQUENCE 261 AA; 29717 MW; 3B89B988605B3639 CRC64;

Query Match
Best Local Similarity 5.7%; Score 91.5; DB 1; Length 261;
Matches 53; Conservative 32; Mismatches 80; Indels 69; Gaps 9;

QY 12 EKKPTGLVTRQS---LEDFPEWERC-EKPLRLHVTYEGTIEEN-GQGMLOVDFANRF 66
Db 48 EKKPYCNAHYPKQSFTMVADTPENLRKQSELSQVRYKEEFKNKGK----- 97

QY 67 VGGVTSAGLVQEEIRFLINPELIIISRLFTVLDHNECLIIITGTEQYSEYTGAYETRWS 126
Db 98 -----FSVADTPELQRIKKTQDIQISNIKYHEEFKS 129

QY 127 RSHEDGSRDDCERRCTEIVADALHFRRLDQFVP-----EKMRELNKAYCG 175
Db 130 RMGPGSGEGNEPERRDSQ---DGSSVRRPLEQQPHIPTSPVYQQPQQPVAQSYGG 185

QY 176 FLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDS 229
Db 186 YKEP-AAPVSIQRSAPG-----GGGKRYRAVDY---SAADEDEVSFDQDGT 228

RESULT 2
K2M2_SHEEP
ID K2M2_SHEEP STANDARD; PRT; 491 AA.
AC P15241;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II microfibrillar, component 7C.
OS Ovis aries (Sheep).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RX MEDLINE=90026244; PubMed=2803231;
RA Sparrow L.G., Robinson C.P., McMahon D.T.W., Rubira M.R.;
RT "The amino acid sequence of component 7C, a type II intermediate-
RL Biochem. J. 261:1015-1022(1989).
CC -!- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -!- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,
CC 7A, 7B, AND 7C).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT MOD_RES 1 1 BLOCKED.
FT DOMAIN 1 109 HEAD.
FT DOMAIN 110 416 ROD.
FT DOMAIN 417 491 TAIL.
FT DOMAIN 110 144 COIL 1A.
FT DOMAIN 145 154 LINKER 1.
FT DOMAIN 155 255 COIL 1B.
FT DOMAIN 256 272 LINKER 12.
FT DOMAIN 273 416 COIL 2.
FT VARIANT 74 74 C -> G OR S.
FT VARIANT 80 80 C -> S.
FT VARIANT 144 144 F -> Y.
FT VARIANT 232 232 S -> V.
FT VARIANT 276 276 C -> D OR N.
FT VARIANT 284 284 Q -> H.
FT UNSURE 1 2 CG -> GC.
SQ SEQUENCE 491 AA; 53681 MW; A801771FE3831ABE CRC64;

Query Match
Best Local Similarity 5.7%; Score 91.5; DB 1; Length 491;
Matches 62; Conservative 32; Mismatches 96; Indels 69; Gaps 13;

QY 50 IEENGQGLQ-VDFANFVGGVTSAGLVQEEIRFL---INPELIIISRLFTVLDHNECL 105
Db 228 LEANSEALIQEIDFLRR-----LYQEEIRVLQANISDTSVIVKMDNSRDLNMDCI 277

QY 106 IITGTEQYSEYT--GYAETVRSRSHEDGSRDDCERRCTEIVADALH---FRRLDQF 160
Db 278 VAEIKAQYDDIASRSRAEAEWYRS-----KCEIKATVIRHGETLRRRTKEEI 325

QY 161 -----VPEKMRRELNKAYC--GFLRPGVS--SENLSAVATGNWCGAFGGDARLKALIQIL 212
Db 326 NELNRVQLRTAEVENAKCQNSKLEAAVTAQEQQGEVALNDARCKLAGLEALQKAKQDM 385

QY 213 AAAAAERDVVYFTFGDSELMRDIIYSMHIFLTERKLTGVGVKLLRLRYNEECNCSTPGP 272
Db 386 ACLLKE-----YQEVMSKLGIDIE----IATYRRLLEG-----EEQRLCEGVG- 425

QY 273 DIKLYPFIYHAVESCAETA 291
Db 426 -----AVNVCVSSS 434

RESULT 3
LASI_MOUSE

```



CC -!- FUNCTION: One of the essential components for the initiation of  
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.  
CC Also involved in the hydrolysis of GTP during the formation of the  
CC 70S ribosomal complex (By similarity).  
CC -!- SUBUNIT: Monomer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).  
CC -!- SIMILARITY: Belongs to the IF-2 family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; BC016590; AAH16590.1; --  
CC MGD; MGI:1924034; Mtif2.  
CC InterPro; IPR000795; EF\_GTPbind.  
CC InterPro; IPR004161; EFTU\_D2.  
CC InterPro; IPR000178; IF2.  
CC InterPro; IPR005225; Small\_GTP.  
CC InterPro; IPR009000; Translat\_factor.  
CC Pfam; PF00009; GTP\_EFTU; 1.  
CC Pfam; PF03144; GTP\_EFTU\_D2; 1.  
CC ProDom; PD186100; IF2; 1.  
CC TIGRFAMs; TIGR00487; IF-2; 1.  
CC TIGRFAMs; TIGR00231; small\_GTP; 1.  
CC PROSITE; PS01176; IF2; 1.  
KW Initiation factor; Protein biosynthesis; GTP-binding; Transit peptide;  
KW Mitochondrion.  
FT TRANSIT 1 29 MITOCHONDRION (POTENTIAL).  
FT CHAIN 30 727 TRANSLATION INITIATION FACTOR IF-2.  
FT DOMAIN 181 330 G-DOMAIN.  
FT NP\_BIND 187 194 GTP (BY SIMILARITY).  
FT NP\_BIND 234 237 GTP (BY SIMILARITY).  
FT NP\_BIND 288 291 GTP (BY SIMILARITY).  
SQ SEQUENCE 727 AA; 81346 MW; A8C3FF8646A45F6D CRC64;  
  
Query Match 5.6%; Score 89; DB 1; Length 727;  
Best Local Similarity 20.7%; Pred. No. 4.4;  
Matches 41; Conservative 34; Mismatches 63; Indels 60; Gaps 10;  
  
QY 121 ETRYWRS-----HEDGSRDDCERRCTEIVADALHF--FRYLDQFVPEK 164  
Db 442 EVIEWRKSEQKEKGKDDLKIMEEGRREHQEAHRKARE--KYGSLHWKERSYI-KFLERK 498  
  
QY 165 MRRELKAYCGFLRPGVSSN---LSAVATGNWCGGAFGGDARLKALIQIL----AAAA 216  
Db 499 QQRP-----LKPKKVERQSNVLPPIIKGD-----VDGSVEAILNLLDLYDASHE 543  
  
QY 217 AERDVVYFTFGDSELMRDYISMHIFLTERKLTGVDVYKLLRYNNEC-----RNCSTPG 271  
Db 544 CELELVHFGGLD-----ISENDVTFATFTDGVYGFNVEAGSAIQQSAQKG 590  
  
QY 272 PDIKLYPFIYHAVESCAE 289  
Db 591 VKIKLHKLIYHLIEDLQE 608  
  
RESULT 5  
ENPP\_BPP21  
ID\_ENPP\_BPP21 STANDARD; PRT; 153 AA.  
AC P27358;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Endopeptidase (EC 3.4.-.-) (Lysis protein Rz).  
GN RZ.  
OS Bacteriophage P21 (Bacteriophage 21).  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC Lambda-like viruses.

OX NCBI\_TaxID=10711;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91210180; PubMed=2019562;  
RA Bonovich M.T., Young R.;  
RT "Dual start motif in two lambdaoid S genes unrelated to lambda S.";  
RL J. Bacteriol. 173:2897-2905(1991).  
CC -!- FUNCTION: Necessary for host cell lysis. It is believed to code  
CC for an endopeptidase that cleaves the amino-carboxyl cross-link  
CC between the diaminopimelic acid and D-alanine residues in the  
CC murein component of the bacterial cell wall.  
CC -!- SIMILARITY: Belongs to peptidase family U8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M65239; AAA32351.1; --  
CC InterPro; IPR004929; Phage\_lysis.  
CC Pfam; PF03245; Phage\_lysis; 1.  
KW Hydrolase; Protease; Bacteriolytic enzyme.  
SQ SEQUENCE 153 AA; 17325 MW; 3291A79D77AA1EFD CRC64;  
  
Query Match 5.4%; Score 86; DB 1; Length 153;  
Best Local Similarity 23.3%; Pred. No. 1.1;  
Matches 42; Conservative 24; Mismatches 56; Indels 58; Gaps 8;  
  
QY 104 CLITGTEQSEYTGTAETYSRWSRSHEDGSRD-----DCERRCTEIVADALHFR 154  
Db 14 CIIVCLSWAVNHYRDNAITYKAQR---DKNARELTLANRVITDIQMFQDVAALDAKYTK 70  
  
QY 155 RYLDQFVPEKM-----RRELN-KAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARL 205  
Db 71 ELADAKAENDALRDDVAAAGRRRLHIKAVCQSVREATTASGVDNAAS-----PRL 119  
  
QY 206 KALIQILAAAAAERDVVYFTFGDSELMRDYISMHIFLTERKLTGVDVYKLLRYNNEECR 265  
Db 120 -----ADTAERD--YFT-----LRRLITMQKLEGTQKYNEQCR 153  
  
RESULT 6  
ENPP\_LAMB  
ID\_ENPP\_LAMB STANDARD; PRT; 153 AA.  
AC P00726;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Endopeptidase (EC 3.4.-.-) (Lysis protein Rz).  
GN RZ.  
OS Bacteriophage lambda.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC Lambda-like viruses.  
OX NCBI\_TaxID=10710;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83189071; PubMed=6221115;  
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;  
RT "Nucleotide sequence of bacteriophage lambda DNA.";  
RL J. Mol. Biol. 162:729-773(1982).  
CC -!- FUNCTION: Necessary for host cell lysis. It is believed to code  
CC for an endopeptidase that cleaves the amino-carboxyl cross-link  
CC between the diaminopimelic acid and D-alanine residues in the  
CC murein component of the bacterial cell wall.  
CC -!- SIMILARITY: Belongs to peptidase family U8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way







|    |  |
|----|--|
| RA | Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  |
| RA | Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  |
| RA | Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,   |
| RA | Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  |
| RA | Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  |
| RA | Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  |
| RA | Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,   |
| RA | Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  |
| RA | Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  |
| RA | Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  |
| RA | Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,   |
| RA | de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,   |
| RA | Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,   |
| RA | Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  |
| RA | de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  |
| RA | da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,   |
| RA | da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  |
| RA | de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  |
| RA | Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  |
| RA | Zago M.A., Zatz M., Meidanis J., Setubal J.C.;   |
| RT | "The genome sequence of the plant pathogen Xylella fastidiosa.";   |
| RL | Nature 406:151-159(2000).  |
| CC | -!- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.   |
| CC | -!- PATHWAY: First step in the sulfate activation pathway. This  |
| CC | reaction occurs early in the reductive branch of the cysteine  |
| CC | biosynthetic pathway.  |
| CC | -!- SUBUNIT: Heterodimer composed of cysD, the smaller subunit, and  |
| CC | cysN (By similarity).  |
| CC | -!- SIMILARITY: Belongs to the PAPS reductase family. CysD subfamily.  |
| CC | -----  |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration   |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |
| CC | the European Bioinformatics Institute. There are no restrictions on its  |
| CC | use by non-profit institutions as long as its content is in no way   |
| CC | modified and this statement is not removed. Usage by and for commercial  |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/   |
| CC | or send an email to license@isb-sib.ch).   |
| CC | -----  |
| DR | EMBL; AE003980; AAF84309.1; ALT_INIT.  |
| DR | PfR; F82672; F82672.   |
| DR | HAMAP; MF_00064; -; 1.   |
| DR | InterPro; IPR002500; PAPS_reduct.  |
| DR | Pfam; PF01507; PAPS_reduct; 1.   |
| KW | Transferase; Nucleotidytransferase; Cysteine biosynthesis;   |
| KW | Complete proteome.   |
| SQ | SEQUENCE 304 AA; 34886 MW; FF4B1EA01BD81EE2 CRC64;   |
|    | Query Match 5.3%; Score 84.5; DB 1; Length 304;  |
|    | Best Local Similarity 22.6%; Pred. No. 3.7;  |
|    | Matches 62; Conservative 32; Mismatches 97; Indels 83; Gaps 14;  |
| QY | 38 PLTRLHVTYEGTIENGCGMLQVDFANRFVGGSAGLVQEIEIRFLINPELI---ISRL 94<br>  :           :   :   :   :   :   :   :   :   :   :                 |
| Dd | 59 PIPLLLHVDTRWKFRE-----MITFRDR---RVATGV---QLRVHNPEGVAQEIPI 105<br>  :           :   :   :   :   :   :   :   :   :   :                 |
| QY | 95 FTEVLHDNECLITGTGEQYSEYTGAEIYWRSRHEDGSDCERRCTEIVADALHFR 154<br>  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :       |
| Dd | 106 THGAAVHTDVMTQTGLRQALEQGQFDAAI-----GGARRDEKSRAKERV----FSFR 154<br>  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   |
| QY | 155 RYLDOFVPEKMRRLENKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQLAA 214<br>  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : |
| Dd | 155 NAHRWDPKNQRPPELMNVVNARIHPG---ESVRVFPLSNW----- 191<br>  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :               |
| QY | 215 AAAERDY-----VYFTFGDSLEMRDIYSMHIFLTRKLTV--GDVVKLL-LR 258<br>  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :         |
| Dd | 192 --TELDDWLVIYREKIPVVPLYFAAAPPVVERD--GMLILVDDRPLLRPGEVSKLLWVR 247<br>  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : |
| QY | 259 YNNEECRCNSTPDPDKLYPFIIHAVESCAETAD 292<br>  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :                           |
| Dd | 248 FRTLGC-----YP-LTGAVESRAATLE 268<br>  :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :                                 |

RESULT 9

```

YCF2 LOTJA
ID YCF2 LOTJA STANDARD; PRT; 2298 AA.
AC Q9B1K6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 270.7 kDa protein ycf2.
GN YCF2-A AND YCF2-B.
OS Lotus japonicus.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eumetazoa I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
OX NCBI_TaxID=34305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Accession MG-20;
RX MEDLINE=21082929; PubMed=11214967;
RA Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;
RT "Complete structure of the chloroplast genome of a legume, Lotus
RT japonicus."
RL DNA Res. 7:323-330(2000).
CC -!- SIMILARITY: Belongs to the ycf2 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP002983; BAB3238.1; -.
CC EMBL; AP002983; BAB33256.1; -.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR008543; DUF825.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF05695; DUF825; 1.
CC SMART; SM00382; AAA; 1.
CC Chloroplast; Hypothetical protein.
KW SEQUENCE 2298 AA; 270688 MW; 36E9A64CA54DC7D0 CRC64;
SQ
-----
Query Match 5.3%; Score 84; DB 1; Length 2298;
Best Local Similarity 21.0%; Pred. No. 53;
Matches 47; Conservative 38; Mismatches 71; Indels 68; Gaps 10;

QY 2 TLF-----CYPRRVTEK-----KPTGLVTFTROSLEDFPEWERCCKPLTRL- 42
Db 670 TLFYEENESFYRIIKKVVQVTCGNDLEDPKPK-IVVFTSNIMEAVNQYRLIRLIPQ 728
QY 43 HVTYEGTIEENGQMLQVDFANFRVGGVTSAGLVQEEIRFLINPELIISR--LFTFVLD 100
Db 729 YITY-----GYGYIR-NVLNRFI-----QKNRFDNRNFYRIQRYQIENDTLN 769
QY 101 H-----NECLITGTEQYSEYTGAYETRWWSRSHEDGSDERD 136
Db 770 HRTMKYTIHQHFSNLKKKSQKKWFDLSLILSRTSMNRDPNAYRYKWSNGNKNFQEHL 829
QY 137 D---CERRCTEIVAIDALHFRRYLDQFVPEKMRRELNKAYCGFL 177
Db 830 DHFISEQNSRFQVFDRLRLINQYSIDWSEVIDKKDLKSLCFFL 873

RESULT 10
ID_RZPD_ECOLI STANDARD; PRT; 153 AA.
AC P75719; P77084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative R2 endopeptidase from lambdoid prophage DLP12 (EC 3.4.-.-).
GN RZPD OR B0556.

```



QY 130 EDGSRDD-----CERRCT-----EIVADALHFRYLDQFVPEKMRRELNKAY 173  
Db 256 EDGHRDPSPPPCVIEKLCSLHDGLVLEAKGIQKQFWKPYIRKLFPEKLLRGKEENL 315  
QY 174 CGFLRPGVSSNLSAV-----ATGNWGCAP-GGDA--RLKALIQILAAAAAERDV 221  
Db 316 TGFLEPGNFAESFKAIVKAYEYVLTATGSLDERIFLGEDAEVEVGTFRCVSAASGTES- 374  
QY 222 VYFTFGDSELMRDYIMHIFLTERKLTGVGVVYLLRYNNEECRNCSCTPGDPIKLYPFIY 281  
Db 375 -----AERTQMRDILQOHL---DKSKTLRVNPLTGVRYVQENSPCVTP----- 415  
QY 282 HAVESCAETADHSGQRTGT 300  
Db 416 -----VSTATHSLNRLHT 428

RESULT 12  
HTPG CLOAB STANDARD; PRT; 624 AA.  
AC Q97E05;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chapterone protein htpg (Heat shock protein htpg) (High temperature protein G).  
GN HTPG OR CAC3315.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."  
RL J. Bacteriol. 183:4823-4838(2001).  
CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By similarity).  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: Belongs to the heat shock protein 90 family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE007828; AAK81247.1; -.  
DR PIR; D97307; D97307.  
DR HAMAP; MF 00505; -; 1.  
DR InterPro; IPR003594; ATPbind\_ATPase.  
DR InterPro; IPR001404; Hsp90.  
DR Pfam; PF02518; HATPase\_c; 1.  
DR Pfam; PF00183; HSP90; 2.  
DR PRINTS; PR00775; HEATSHOCK90.  
DR SMART; SM00387; HATPase\_c; 1.  
DR PROSITE; PS00298; HSP90; 1.  
KW Chaperone; ATP-binding; Heat shock; Complete proteome.  
FT DOMAIN 1 341 A; SUBSTRATE-BINDING (BY SIMILARITY).  
FT DOMAIN 342 550 B (BY SIMILARITY).  
FT DOMAIN 551 624 C.  
SQ SEQUENCE 624 AA; 72407 MW; 131C769B501EB26C CRC64;

Query Match 5.2%; Score 82.5; DB 1; Length 624;  
Best Local Similarity 21.9%; Pred. No. 14;  
Matches 43; Conservative 25; Mismatches 71; Indels 57; Gaps 10;  
QY 5 CYFRRVTEKKPTGLVTFTRQSLDFFEWERCCKPLRLHVTYEG---TIEENGQGMQLQVD 61  
Db 43 CYRSLVDTN-----ITFNK---DDF-----YIRISADKENKTLTITDTGIGMTKDD 86  
QY 62 FANRFVGGVTSAGLVQEEIRFLINPELLIISRLFTTEVLHDHNECLITGTEQYSEYTG-- 119  
Db 87 LENNL--GTIAKSG-----SFAKSENEAKEGVLDIGQFGVGFSAFMV 128  
QY 120 -----AETRWRSRSHEDGSRDDCERRC--TEIV--AIDALHFRYLDQFVP 162  
Db 129 ADDVTVISRSVDSEAYKWESKGVGTIEKCEKETPGTEIVLKIKENTDDEKY-DEFLD 187  
QY 163 EKMRRELNKAYCGFLR 178  
Db 188 EYKLRSLIKYSDFIK 203  
RESULT 13  
AKA3 BOVIN STANDARD; PRT; 854 AA.  
ID AKA3 BOVIN  
AC O77797;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE A-kinase anchor protein 3 (Protein kinase A anchoring protein 3) (PKA3) (A-kinase anchor protein 110 kDa) (AKAP 110) (Fragment).  
GN AKAP3 OR AKAP110.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=99252879; PubMed=10319321;  
RA Vijayaraghavan S., Liberty G.A., Mohan J., Winfrey V.P., Olson G.E., Carr D.W.;  
RT "Isolation and molecular characterization of AKAP110, a novel, sperm-specific protein kinase A-anchoring protein."  
RL Mol. Endocrinol. 13:705-717(1999).  
CC -!- FUNCTION: May function as a regulator of both motility- and head-associated functions such as capacitation and the acrosome reaction (By similarity).  
CC -!- SUBCELLULAR LOCATION: Localizes to the ribs of the fibrous sheath in the principal piece of the sperm tail (By similarity).  
CC -!- DOMAIN: RII-binding site, predicted to form an amphipathic helix, could participate in protein-protein interactions with a complementary surface on the R-subunit dimer.  
CC -!- PTM: Phosphorylated on tyrosine (By similarity).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF093407; AAC63370.1; -.  
DR InterPro; IPR008382; AKAP\_110.  
DR Pfam; PF05716; AKAP\_110; 1.  
KW Phosphorylation.  
FT NON TER 1 140 PKA-RII SUBUNIT BINDING DOMAIN.  
FT DOMAIN 127 140  
SQ SEQUENCE 854 AA; 94216 MW; 0AB67B6E0327A32B CRC64;  
Query Match 5.2%; Score 82.5; DB 1; Length 854;  
Best Local Similarity 23.0%; Pred. No. 20;







```
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=Long;
CC IsoId=Q13191-1; Sequence=Displayed;
CC Name=Truncated 1;
CC IsoId=Q13191-2; Sequence=VSP_005729;
CC Name=Truncated 2;
CC IsoId=Q13191-3; Sequence=VSP_005730, VSP_005731;
CC -!- TISSUE SPECIFICITY: Expressed in lung, kidney, spleen, and testis,
CC as well as fetal brain and liver and hematopoietic cell lines, but
CC not in adult brain, liver, pancreas, salivary gland, or skeletal
CC muscle.
CC -!- DOMAIN: The N-terminus is composed of the phosphotyrosine binding
CC (PTB) domain, a short linker region and the RING-type zinc finger.
CC The PTB domain, which is also called TKB (tyrosine kinase binding)
CC domain, is composed of three different subdomains: a four-helix
CC bundle (4H), a calcium-binding EF hand and a divergent SH2 domain
CC (By similarity).
CC -!- DOMAIN: The RING-type zinc finger domain mediates binding to an E2
CC ubiquitin-conjugating enzyme (By similarity).
CC -!- PTM: Phosphorylated on tyrosine residues.
CC -!- MISCELLANEOUS: This protein has one functional calcium-binding
CC site (By similarity).
CC -!- SIMILARITY: Contains 2 EF-hand-like calcium-binding domains.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: Contains 1 SH2 domain.
CC -!- SIMILARITY: Contains 1 UBA domain.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chroncancer/Genes/CBLbID193.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U26710; AAB09291.1; -
CC EMBL; U26711; AAB09292.1; -
CC EMBL; U26712; AAB09293.1; -
CC HSSP; P22681; 1B47.
CC Genew; HGNC:1542; CBLB.
CC MIM; 604491; -
CC GO; GO:0005515; F:protein binding; TAS.
CC GO; GO:0008270; F:zinc ion binding; TAS.
CC GO; GO:0006507; P:NLS-bearing substrate-nucleus import; TAS.
CC GO; GO:0007165; P:signal transduction; TAS.
CC InterPro; IPR003153; Cbl_N.
CC InterPro; IPR000980; SH2.
CC InterPro; IPR000449; UBA domain.
CC InterPro; IPR001841; Znf_ring.
CC Pfam; PF02761; Cbl_N2; 1.
CC Pfam; PF02762; Cbl_N3; 1.
CC Pfam; PF02262; Cbl_N; 1.
CC Pfam; PF00627; UBA; 1.
CC Pfam; PF00097; zf-C3HC4; 1.
CC SMART; SM00184; RING; 1.
CC SMART; SM00252; SH2; 1.
CC SMART; SM00165; UBA; 1.
CC PROSITE; PS50030; UBA; 1.
CC PROSITE; PS00518; ZF_RING_1; 1.
CC PROSITE; PS50089; ZF_RING_2; 1.
CC Nuclear protein; Zinc-finger; Phosphorylation; Alternative splicing;
KW SH2 domain; Calcium-binding.
FT DOMAIN 36 349 PTB.
FT DOMAIN 36 169 4H.
FT DOMAIN 204 212 EF-HAND LIKE 1.
FT CA BIND 221 232 EF-HAND LIKE 2.
FT DOMAIN 259 333 SH2.
FT DOMAIN 334 372 LINKER.
FT ZN FING 373 412 RING-TYPE.
FT DOMAIN 477 701 PRO-RICH.
```

|  |   |         |            |  |
|--|---|---------|------------|--|
| FT   | DOMAIN  | 931     | 970        | UBA.                                     |
| FT   | DOMAIN  | 116     | 119        | NUCLEAR LOCALIZATION SIGNAL (POTENTIAL). |
| FT   | BINDING   | 286     | 286        | PHOSPHOTYROSINE (BY SIMILARITY).         |
| FT   | VARSPLIC  | 812     | 982        | Missing (in isoform Truncated 1).        |
| FT   |   |         |            | /FTid=VSP_005729.                        |
| FT   | VARSPLIC  | 767     | 770        | DVFD -> TYRI (in isoform Truncated 2).   |
| FT   |   |         |            | /FTid=VSP_005730.                        |
| FT   | VARSPLIC  | 771     | 982        | Missing (in isoform Truncated 2).        |
| FT   |   |         |            | /FTid=VSP_005731.                        |
| SQ   | SEQUENCE  | 982 AA; | 109479 MW; | 038D4965DAFB98CB CRC64;                  |
| Query Match 5.1%; Score 81.5; DB 1; Length 982;                  |   |         |            |  |
| Best Local Similarity 18.2%; Pred.No. 30;                        |   |         |            |  |
| Matches 59; Conservative 52; Mismatches 133; Indels 81; Gaps 15; |   |         |            |  |
| QY   | 6 YFRRVTEKKPTGLVTFTTRQSLDEPWEKCEKPLTRLHVHYEGTTEENGQGMQLQVDFAN-  | 64      |            |  |
| Db   | 110 YIDSLMKSKRAIRLFKEGKERMVEEQSDRRNLTKLSLIFSHMLAE-----IKAI      | 164     |            |  |
| QY   | 65 RFVGGG--VTSAGLVQEEIRFLINPELISRLFTFVLDHNECLIIITGEQYSEYTG      | 122     |            |  |
| Db   | 165 QFQGDNFRITKADAAEFWRKFFGDKTI                                 | 204     |            |  |
| QY   | 123 YRWSRSHEDGSRDDCERRCTEIVADALHFRRYLDQFVP-EKMRRELN-----KAYCG   | 175     |            |  |
| Db   | 205 HQISSSLEAMALKSTIDLTCDYISV--FEEDIFTRLFQPWGSILRNWNFLAVTHPGYMA | 262     |            |  |
| QY   | 176 FL-----RPGVSSENLSAVATGNWCGGAFGGDARL-----KALIQILAA           | 214     |            |  |
| Db   | 263 FLTYDEVKARLQYSTKPGSYIFRLSCTRLGOWAIGYVTDGNILQTIPIHNKPLFQALID | 322     |            |  |
| QY   | 215 AAAERDVVYFTFGDSELMR-DIYSM-----HIFLTERK-----LTVGDVYKLLRLRYNE | 262     |            |  |
| Db   | 323 GSREG---FLLYPDGRSYNPDLTGLCEPTPHDIKVTQEQYELYCEMGSTFQL-----   | 372     |            |  |
| QY   | 263 ECRNCSTPGPDIKLYPFIYHAVESC                                   | 287     |            |  |
| Db   | 373 -CKICAENDKDVKEPCGHLMCTSC                                    | 396     |            |  |

Search completed: May 26, 2004, 18:41:11  
Job time : 6.80653 secs



Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGIA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 916  
QY 241 FLTERKLTVDVYKLLRLRYNNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTVDVYKLLRLRYNNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976

RESULT 2

Q86W56 PRELIMINARY; PRT; 976 AA.  
ID Q86W56  
AC Q86W56;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;

Query Match 99.1%; Score 1583; DB 4; Length 976;

Best Local Similarity 99.3%; Pred. No. 3.1e-143;

Matches 298; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTQSLDFPEWERCCKPLTRLHVYEGTIEENGQGMLOV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTQSLDFPEWERCCKPLTRLHVYEGTIEENGQGMLOV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGIA 120  
Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGIA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 916  
QY 241 FLTERKLTVDVYKLLRLRYNNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTVDVYKLLRLRYNNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976

RESULT 3

Q7Z742 PRELIMINARY; PRT; 976 AA.  
ID Q7Z742  
AC Q7Z742;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC052966; AAH52966.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;

Query Match 98.8%; Score 1578; DB 4; Length 976;

Best Local Similarity 99.0%; Pred. No. 9.4e-143;

Matches 297; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTLFCYFRRVTEKKPTGLVTFTQSLDFPEWERCCKPLTRLHVYEGTIEENGQGMLOV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTQSLDFPEWERCCKPLTRLHVYEGTIEENGQGMLOV 736  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGIA 120  
Db 737 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNECLIIITGTEQYSEYTGIA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVAIDALHFRRYLDQFVPEKMRRELKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 240  
Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHI 916  
QY 241 FLTERKLTVDVYKLLRLRYNNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 917 FLTERKLTVDVYKLLRLRYNNEECRNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976

RESULT 4

O88622 PRELIMINARY; PRT; 968 AA.  
ID O88622  
AC O88622;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99380098; PubMed=10449915;  
RA Ame J.C.; Apio F.; Jacobson E.L.; Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
RT hybridization."  
RL Cytogenet. Cell Genet. 85:269-270 (1999).  
DR EMBL; AF079557; AAC28735.1; --  
DR MGD; MGI:1347094; Parg.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;  
  
Query Match 93.7%; Score 1497; DB 11; Length 968;  
Best Local Similarity 93.7%; Pred. No. 5.9e-135;  
Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;  
  
QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 60  
Db 669 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 728  
  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQYSEYTGIA 120  
Db 729 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQYSEYTGIA 788  
  
QY 121 ETYRWSRSHEDGSRDCCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 789 ETYRWSRSHEDGSRDCCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 848  
  
QY 181 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMDIYSMHI 240  
Db 849 VPSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMDIYSMHI 908  
  
QY 241 FLTERKLTVDGVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 300  
Db 909 FLTERKLDVGKVKYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 968  
  
RESULT 5  
O02776 PRELIMINARY; PRT; 977 AA.  
AC O02776;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN BPARG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9727328; PubMed=9115250;  
RA Lin W.; Ame J.C.; Aboul-Ela N.; Jacobson E.L.; Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding bovine poly (ADP-ribose) glycohydrolase."  
RT J. Biol. Chem. 272:11895-11901 (1997).  
RL EMBL; U78975; AAB53370.1; --  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;  
  
Query Match 93.5%; Score 1493; DB 6; Length 977;  
Best Local Similarity 94.0%; Pred. No. 1.4e-134;

Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;  
  
QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 60  
Db 678 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 737  
  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQYSEYTGIA 120  
Db 738 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQYSEYTGIA 797  
  
QY 121 ETYRWSRSHEDGSRDCCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 798 ETYRWSRSHEDGSRDCCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 857  
  
QY 181 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMDIYSMHI 240  
Db 858 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMDIYSMHI 917  
  
QY 241 FLTERKLTVDGVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 299  
Db 918 FLTERKLTVDGVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHAVESCAETADHSGQRTGT 976  
  
RESULT 6  
Q80YQ6 PRELIMINARY; PRT; 961 AA.  
AC Q80YQ6;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; BC050892; AAH50892.1; --  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;  
  
Query Match 92.4%; Score 1476; DB 11; Length 961;  
Best Local Similarity 94.9%; Pred. No. 6.1e-133;  
Matches 277; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
  
QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 60  
Db 670 KTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHLVHYEGTIEENGQGMLOV 729  
  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQYSEYTGIA 120  
Db 730 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEQYSEYTGIA 789  
  
QY 121 ETYRWSRSHEDGSRDCCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 790 ETYRWSRSHEDGSRDCCERRCTEIVAIDALHFRYLDQFVPEKMRRELKAYCGFLRPG 849  
  
QY 181 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMDIYSMHI 240  
Db 850 VPSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMDIYSMHI 909  
  
QY 241 FLTERKLTVDGVYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHAVESCAETAD 292  
Db 910 FLTERKLDVGKVKYKLLRLRYNEECRCNCSTPGPDIKLYPFIYHAVESCAETAD 961  
  
RESULT 7



Q9QYM2 Q9QYM2 PRELIMINARY; PRT; 972 AA.  
AC Q9QYM2;  
DT Q9QYM2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN PARG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BUF; TISSUE=Colon;  
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,  
RA Aoki Y., Nakagawa H., Sugimura T.;  
RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase  
RT (Parg).";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB019366; BAA87901.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;  
  
Query Match 91.7%; Score 1464; DB 11; Length 972;  
Best Local Similarity 91.0%; Pred. No. 8.9e-132;  
Matches 273; Conservative 12; Mismatches 15; Indels 0; Gaps 0;  
  
QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLDFPEWERCCKPLTRLHLVTEGTEIENGQGMLOV 60  
Db 673 KTLFCYFRRVTEKKPTGLVTFTRQSLDFPEWERCCKPLTRLHLVTEGTEIENGQGMLOV 732  
  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLTEVLDHNECLIIITGTEQYSEYTGIA 120  
Db 733 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLTEVLDHNECLIIITGTEQYSEYTGIA 792  
  
QY 121 ETYRWSRSHEDGSEKDDQRCCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 793 ETYRWSRSHEDGSEKDDQRCCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 852  
  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIIQLAAAAAERDVVYFTFGDSELMRDIIYSMHI 240  
Db 853 VPPENLSAVATGNWCGAFGGDARLKALIIQLAAAAAERDVVYFTFGDSELMRDIIYSMHI 912  
  
QY 241 FLTERKLTGVDVYKLLRYNEECRNCTPGPDIKLYPIYHVESCAETADHSQRTGT 300  
Db 913 FLTERKLTGVDVYKLLRYNEECRNCTPGPDIKLYPIYHVESCAETADHSQRTGT 972  
  
RESULT 8  
Q8CB72 Q8CB72 PRELIMINARY; PRT; 920 AA.  
AC Q8CB72;  
DT Q8CB72;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium;  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).  
DR EMBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;  
  
Query Match 78.9%; Score 1260; DB 11; Length 920;  
Best Local Similarity 94.4%; Pred. No. 3.7e-112;  
Matches 237; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
  
QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLDFPEWERCCKPLTRLHLVTEGTEIENGQGMLOV 60  
Db 670 KTLFCYFRRVTEKKPTGLVTFTRQSLDFPEWERCCKPLTRLHLVTEGTEIENGQGMLOV 729  
  
QY 61 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLTEVLDHNECLIIITGTEQYSEYTGIA 120  
Db 730 DFANRFVGGVTSAGLVQEEIRFLINPELIIISRLTEVLDHNECLIIITGTEQYSEYTGIA 789  
  
QY 121 ETYRWSRSHEDGSEKDDQRCCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
Db 790 ETYRWSRSHEDGSEKDDQRCCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 849  
  
QY 181 VSSENLSAVATGNWCGAFGGDARLKALIIQLAAAAAERDVVYFTFGDSELMRDIIYSMHI 240  
Db 850 VPSENLSAVATGNWCGAFGGDARLKALIIQLAAAAAERDVVYFTFGDSELMRDIIYSMHI 909  
  
QY 241 FLTERKLTGVD 251  
Db 910 FLTERKLTGVD 920  
  
RESULT 9  
Q960N8 Q960N8 PRELIMINARY; PRT; 723 AA.  
AC Q960N8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE LD42380p.  
GN PARG OR EG:114E2.1 OR CG2864.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY051955; AAK93379.1; -.  
DR FlyBase; FBgn0023216; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;  
  
Query Match 38.8%; Score 619.5; DB 5; Length 723;  
Best Local Similarity 46.4%; Pred. No. 1.3e-50;  
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;  
  
QY 1 KTLFCYFRRV--TEKK---PTGLVTFTRQS--LEDFPEWERCCKPL--TELHVTYEGTI 50  
Db 252 KCIMHYFRRVCPTEADSNVPTGVVTFRRSGLPHELDWSQSAAPGLDVLHVDAGETI 311  
  
QY 51 HENGQGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLTEVLDHNECLIIITGT 110  
Db 312 EDEGIGLLQVDFANKYLGGLGGVGLGHCVCQEEIRFVICPELLVGLKFTCELRPEALVMLGA 371











GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:28:37 ; Search time 162.413 Seconds  
(without alignments)  
13270.296 Million cell updates/sec

Title: US-09-302-812-5  
Perfect score: 6773  
Sequence: 1 gggggactgtgtgctgcggg.....aatcattgtcagaaaaaa 3814

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues  
Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO spool\_p/6333148/runat\_26052004\_150052\_5654/app\_query.fasta\_1.12437  
-DB=A\_Geneseq\_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148 @CGN 1 1 414 @runat\_26052004\_150052\_5654 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID         | Description        |
|------------|--------|-------------|--------|------------|--------------------|
| 1          | 5164   | 76.2        | 968    | 5 AAE25631 | Aae25631 Murine po |
| 2          | 5164   | 76.2        | 968    | 5 AAU76022 | Aau76022 Mouse pol |
| 3          | 5164   | 76.2        | 968    | 5 AAU76013 | Aau76013 Mouse pol |
| 4          | 5164   | 76.2        | 968    | 6 ABG72280 | Abg72280 Murine p  |
| 5          | 4395   | 64.9        | 976    | 5 AAE25630 | Aae25630 Human pol |
| 6          | 4395   | 64.9        | 976    | 5 AAU76021 | Aau76021 Human pol |
| 7          | 4395   | 64.9        | 976    | 5 AAU76012 | Aau76012 Human pol |
| 8          | 4395   | 64.9        | 976    | 6 ABG72279 | Abg72279 Human pol |
| 9          | 4332.5 | 64.0        | 977    | 5 AAE25629 | Aae25629 Bovine po |
| 10         | 4332.5 | 64.0        | 977    | 5 AAU76020 | Aau76020 Bovine po |

|    |        |      |      |            |                     |
|----|--------|------|------|------------|---------------------|
| 11 | 4332.5 | 64.0 | 977  | 5 AAU75799 | Aau75799 Bovine po  |
| 12 | 4332.5 | 64.0 | 977  | 6 ABG72278 | Abg72278 Bovine po  |
| 13 | 1074.5 | 15.9 | 768  | 4 ABB59491 | Abb59491 Drosophil  |
| 14 | 1074.5 | 15.9 | 768  | 5 AAE25632 | Aae25632 Fruit fly  |
| 15 | 1074.5 | 15.9 | 768  | 5 AAU76023 | Aau76023 Fruit fly  |
| 16 | 1074.5 | 15.9 | 768  | 5 AAU76014 | Aau76014 Fruit fly  |
| 17 | 1074.5 | 15.9 | 768  | 6 ABG72281 | Abg72281 Fruit fly  |
| 18 | 443.5  | 6.5  | 726  | 5 AAE25633 | Aae25633 Poly aden  |
| 19 | 443.5  | 6.5  | 726  | 5 AAU76024 | Aau76024 Worm poly  |
| 20 | 443.5  | 6.5  | 726  | 5 AAU76015 | Aau76015 Worm poly  |
| 21 | 443.5  | 6.5  | 726  | 6 ABG72282 | Abg72282 C. elegan  |
| 22 | 350.5  | 5.2  | 819  | 4 ABG20721 | Abg20721 Novel hum  |
| 23 | 199    | 2.9  | 100  | 4 ABG11103 | Abg11103 Novel hum  |
| 24 | 190    | 2.8  | 931  | 4 ABB61093 | Abb61093 Drosophil  |
| 25 | 186    | 2.7  | 33   | 5 AAE25652 | Aae25652 Bovine po  |
| 26 | 178    | 2.6  | 2703 | 4 ABB60074 | Abb60074 Drosophil  |
| 27 | 177.5  | 2.6  | 2951 | 4 ABB60291 | Abb60291 Drosophil  |
| 28 | 174    | 2.6  | 718  | 4 AAM80140 | Aam80140 Human pro  |
| 29 | 172    | 2.5  | 1489 | 6 ABO14779 | Abol14779 Novel hum |
| 30 | 172    | 2.5  | 1489 | 6 ABU11038 | Abu11038 Human pro  |
| 31 | 169.5  | 2.5  | 1297 | 6 ABU43768 | Abu43768 Protein e  |
| 32 | 169    | 2.5  | 699  | 4 AAM79156 | Aam79156 Human pro  |
| 33 | 168    | 2.5  | 335  | 4 ABG09631 | Abg09631 Novel hum  |
| 34 | 167    | 2.5  | 31   | 5 AAE25634 | Aae25634 Bovine po  |
| 35 | 167    | 2.5  | 31   | 5 AAE25651 | Aae25651 Bovine po  |
| 36 | 167    | 2.5  | 31   | 5 AAU76025 | Aau76025 Bovine po  |
| 37 | 167    | 2.5  | 31   | 5 AAU76016 | Aau76016 Bovine po  |
| 38 | 167    | 2.5  | 31   | 6 ABG72283 | Abg72283 Oligopept  |
| 39 | 167    | 2.5  | 1169 | 6 ABU43255 | Abu43255 Protein e  |
| 40 | 167    | 2.5  | 1177 | 5 ABP40261 | Abp40261 Staphyloc  |
| 41 | 166.5  | 2.5  | 688  | 4 ABG03434 | Abg03434 Novel hum  |
| 42 | 166.5  | 2.5  | 787  | 6 ABP55414 | Abp55414 Human MDD  |
| 43 | 166    | 2.5  | 1187 | 2 AAR66451 | Aar66451 AF-4 prot  |
| 44 | 166    | 2.5  | 1210 | 2 AAR66450 | Aar66450 AF-4 prot  |
| 45 | 166    | 2.5  | 1210 | 7 ABR63878 | Abrr63878 Human AF4 |

ALIGNMENTS

RESULT 1  
AAE25631  
ID AAE25631 standard; protein; 968 AA.  
XX  
AC AAE25631;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
XX  
KW Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Mus musculus.  
XX  
PN US6395543-B1.  
XX  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
DR WPI; 2002-535641/57.  
DR N-PSDB; AAD42083.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
PS Claim 3; Col 63-68; 77pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
CC The present sequence is murine PARG  
XX  
SQ Sequence 968 AA;

Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 5164.00 Matches: 968  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 76.24% Indels: 0  
DB: 5 Gaps: 0

US-09-302-812-5 (1-3814) x AAE25631 (1-968)

QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCAGAAAGCGCGTGGGCGCGCTGGAAC 87  
Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20  
QY 88 TCTGCGCGGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGCGGTCTCGAC 147  
Db 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeuAsp 40  
QY 148 CCCAAGGACGCTCCCGTCCAGTTCAGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCGGG 207  
Db 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
QY 208 CGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTCCTTCAACAAAGACTTACT 267  
Db 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
QY 268 ACTTGGATGGATACTAAAGGACCCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAT 327  
Db 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
QY 328 ACAAGAATTGACTCCATGATGAGTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTG 387  
Db 101 ThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyrProHisLysVal 120  
QY 388 GAAAAATTGAAAAATGTTCCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGTTCA 447  
Db 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer 140  
QY 448 CAGTATTTGAACCAACAGCAGACTGCGAGTGTGTCAAGTGGCAGAAATGAAGGGAGCAT 507  
Db 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160  
QY 508 GCAGAACAGCTTTTGGCAAGTGAAGCCTCCCGGGGACTCCGCTACCAAGCAGCTTAGT 567  
Db 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
QY 568 AATGCTAACATTGGTCAGTACCCACACTGATGACCACAGTGACACAGATCATGAAGAA 627  
Db 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200

QY 628 GACAGAGACAATCAGCAGTTTCTTACACCTATATAAACTTGCAAATACAAAGCCACAGTA 687  
Db 201 AspArgAspAsnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220  
QY 688 GGAGATGGCAGGCCAGAACAACTGTAAGTGCAGTGGATCTGCCAGTCTGTGAAAGAC 747  
Db 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240  
QY 748 TGTACAGGCTGTCAACAGGAGGAGGTGGATGTCTTACCAGAGAGTCTCTTGTACAGATGT 807  
Db 241 CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal 260  
QY 808 GGTCCGAGGACATTGGAACCTGGACCAAAAAATGACAAATGACTGGACAAGAAC 867  
Db 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280  
QY 868 AGCTAGGTGATTGCGCTCCATTGAGAAAGAAAGTGAAGCCTGAGTCACCAATGGATGTA 927  
Db 281 SerLeuGlyAspSerProPropheGluLysGluSerGluProGluSerProMetAspVal 300  
QY 928 GACAACTCGAGAAACAGTTGTCAAGATTTCAGAGCAGATGAAGAAACAAGTCCAGTCTTT 987  
Db 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
QY 988 GATGAGCAAGATGATCGTTCTCCAAAACAGCAATAAACTTTCAAGTTGCCAAGCAAGA 1047  
Db 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerCysGlnAlaArg 340  
QY 1048 GAAAGCTGATGGCGATCTTAGGAAACGGTATTTTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107  
Db 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
QY 1108 TTCCAATTTGAAGGAGAAAATAATGTCTGGGACCAGTGAATTAATGCCAAGCCATCTTGA 1167  
Db 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAGGGATTCT 1227  
Db 381 AsnSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
QY 1228 AAAATTACAGATCATTTTCATGAGAATTTCCAAAGTCAGAGGACAGAGAAAGAAACAATGT 1287  
Db 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGlnCys 420  
QY 1288 GAAAGTCAGACATCAAGAAACAGAAAGGAAGATTCCAAAATACATCCCACCTACCTCCCT 1347  
Db 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440  
QY 1348 CCAGAGAAGAAAGTGGCTGGGAACTCCTATTGAGGAATGAGAAAATGCCGTGGTGGG 1407  
Db 441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460  
QY 1408 ATCCATTGCTTCCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTT 1467  
Db 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
QY 1468 CTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTGTGGGATAAC 1527  
Db 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500  
QY 1528 AAACATGTGAAAATGCCTTGTTCGGAAACAAAATTTGTACCCCTGTGGAGATGAGAATGGT 1587  
Db 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
QY 1588 GAGCGAACTGCAGGAGTAGGTGGGAGCTATTTCAGACTGCACCTTCTCAACAAATTCACA 1647  
Db 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540  
QY 1648 CGACCCCGAAGTGAAGGATGCGATTCTGAAATACAAATGTGGCATATTCTAAGAAATGG 1707  
Db 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560

QY 1708 GACTTTACAGCTTTGGTTGATTTCTGGATAAGGTACTTGAAGAAGCAGAGGCCCAACAT 1767  
Db 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGlnHis 580  
QY 1768 TTATATCAGTCCATTTTACCTGACATGGTGAAAAATTGCACTCTGTCTGCCAAATATTTC 1827  
Db 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCCAGCCCAATACCACCTCCTGAAACAGAAGATGAATCAATTCTGTACGATGTACAGGAA 1887  
Db 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATCGCCAGTCTTTTAGTAATGCTTTCTTCTGCAATTTCCCGACGGAATGCCAAG 1947  
Db 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640  
QY 1948 ATGAAATCGGAGTATCTAGTTACCCAGACATTAACCTCAATCGGTTGTTTGAAGACGT 2007  
Db 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAAGGAAACCCAGAAAACACTGAAAACACTCTTCTGCTACTTTTCGAAGAGTCACAGAG 2067  
Db 661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu 680  
QY 2068 AAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGTCTTTGAAGATTTTCCAGAATGG 2127  
Db 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp 700  
QY 2128 GAAAGGTGTGAAAAGCCCTCTGACACGCTTACAGCTACTTACGAGGGTACCATAGAAGGC 2187  
Db 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGGCCGAGGCATGTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGTGTGACT 2247  
Db 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGGACTTGTACAGAAGAAATCAGATTTTAAATCAATCCTGAATTGATTTTCA 2307  
Db 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGCTGTTCACTGAGGTGTGGATCACAAATGAGTGTCTTATTATCACAGGTACTGAACAG 2367  
Db 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCCGAAGCCATGAAGATGGG 2427  
Db 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAAAGGACGATTGGCAGCGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCAC 2487  
Db 801 SerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCT 2547  
Db 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840  
QY 2548 TACTGCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTCTGCAGTGGCCACGGGA 2607  
Db 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGCTGTGGTGCCTTTGGGGTGACGCTAGATTAAAGCCCTTAATACAGATCCTG 2667  
Db 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
QY 2668 GCAGCTGCTGGGCTGAACGTGACGTGGTGTATTATTTACCTTTGGGACTCAGAGTTGATG 2727  
Db 881 AlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTTGAAAAGTG 2787  
Db 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
QY 2788 TACAAGTTATTGCTTAGATACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA 2847

Db 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTATCCATTCATATACCATGCTGTGAGTCAAGTGCAGAGACCACGTGAC 2907  
Db 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerAlaGluThrThrAsp 960  
QY 2908 ATGCCAGGACAGAGGCGAGCCACC 2931  
Db 961 MetProGlyGlnLysAlaGlyThr 968  
RESULT 2  
AAU76022  
ID AAU76022 standard; protein; 968 AA.  
XX  
AC AAU76022;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
OS Mus musculus.  
XX  
PN US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
PI WPI; 2002-163240/21.  
XX  
DR N-PSDB; ABK14933.  
XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
PS Claim 2; Col 63-70; 81pp; English.  
XX  
CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the mouse PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX  
SQ Sequence 968 AA;  
Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 5164.00 Matches: 968



|   |         |   |      |
|---|---------|---|------|
| Percent Similarity:                         | 100.00% | Conservative:   | 0    |
| Best Local Similarity:                      | 100.00% | Mismatches:   | 0    |
| Query Match:                                | 76.24%  | Indels:   | 0    |
| DB:   | 5       | Gaps:   | 0    |
| US-09-302-812-5 (1-3814) x AAU76022 (1-968) |         |   |      |
| QY  | 28      | ATGAGTGGGCCCCCGCTGGAGCCCTGCACGAAAGCGCGCTGGGGCGCGCTGGAAC       | 87   |
| Db  | 1       | MetSerAlaGlyProGlyTyrGluProCysThrLysAlaArgTyrGlyAlaAlaGlyThr  | 20   |
| QY  | 88      | TCTGCGCGACTGCCTCGGACTCCGGAGCTTCCCTGGCAGGCAGAGCGTGTCTCGAC      | 147  |
| Db  | 21      | SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeuAsp  | 40   |
| QY  | 148     | CCCAAGGACGCTCCCGTCCAGTTCAGGTCCCTCCGTCCTCGCCAGCCTCGCTCGGGG     | 207  |
| Db  | 41      | ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly  | 60   |
| QY  | 208     | CGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTGTTTCAAAACAAAGACTTACT    | 267  |
| Db  | 61      | ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr  | 80   |
| QY  | 268     | ACTTGGATGGATACTAAAGGACCCCAAGACAGCTGAATCAGAAAGTAAGAAAACAACAT   | 327  |
| Db  | 81      | ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn     | 100  |
| QY  | 328     | ACAAGAATTGACTCCATGATGAGTTCTGTGTCAGAAAGATAACTTTTACCCACATAAGGTG | 387  |
| Db  | 101     | ThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyrProHisLysVal     | 120  |
| QY  | 388     | GAAAAATTGGAAAAATGTTCTCAGCTAAATCTTGATAAATCACCCACAGAAAGAGTTCA   | 447  |
| Db  | 121     | GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer  | 140  |
| QY  | 448     | CAGTATTTGAACCAACAGCAGACTGCGAGTGTGTCAGTGGCAGAAATGAAGGAGCAT     | 507  |
| Db  | 141     | GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTyrGlnAsnGluGlyLysHis     | 160  |
| QY  | 508     | GCAGAACAGCTTTTGGCAAGTGAGCTCCCGCGGGAGCTCCGCTACCAAGCAGCTTAGT    | 567  |
| Db  | 161     | AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer  | 180  |
| QY  | 568     | AATGCTAACATTGGTCAGTCACCCACACTGATGATCACACAGTGACACAGATCATGAAGAA | 627  |
| Db  | 181     | AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu  | 200  |
| QY  | 628     | GACAGAGACAATCAGCAGTTTCTTACACCTATAAACTTGCAATACAAAGCCACAGTA     | 687  |
| Db  | 201     | AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal  | 220  |
| QY  | 688     | GGAGATGGCGCAGGCCAGAACCACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAGAC   | 747  |
| Db  | 221     | GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp  | 240  |
| QY  | 748     | TGTACAGGCTGTCAACAGGAGGAGGTGGATGTGTCACAGAGAGTCCCTTTGTCAGATGTT  | 807  |
| Db  | 241     | CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal     | 260  |
| QY  | 808     | GGTCCGAGGACATTGGAACCTGGACCAAAAAATGACAAATAATGACTGGACAAAGC      | 867  |
| Db  | 261     | GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer  | 280  |
| QY  | 868     | AGCCTAGGTGATTCCGCTCCATTGTGAGAAAGAAAGTCAGCCTGAGTCACCAATGGATGA  | 927  |
| Db  | 281     | SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal  | 300  |
| QY  | 928     | GACAACTCGAGAAACAGTTGTCAAGATTCAGAACGAGATGAAGAAACAAGTCCAGTCTTT  | 987  |
| Db  | 301     | AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe  | 320  |
| QY  | 988     | GATGAGCAAGATGATCGTTCTCTCCCAACAGCAATAAACTTTCAAGTTGCCAAGCAAGA   | 1047 |

|    |      |   |      |
|----|------|---|------|
| Db | 321  | AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg  | 340  |
| QY | 1048 | GAAGCTGATGGCGATCTTAGGAAACGGTATTTTGACTAAGGGAAGTGAAGTTAGATTGCAT | 1107 |
| Db | 341  | GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis  | 360  |
| QY | 1108 | TTCCAATTTGAAGGAGAAAAATAATGCTGGGACCAGTGAATTAATGCCAAGCCATCTGGA  | 1167 |
| Db | 361  | PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly  | 380  |
| QY | 1168 | AACTCTTCTAGCCTTAATGTAGAGTGTAGAGTTTCCAAGCAGCATGGAAAAAGGATTCT   | 1227 |
| Db | 381  | AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer  | 400  |
| QY | 1228 | AAAATTACAGATCATTTTCATGAGAATTTTCCAAGTCAGAGGACAGAAAGAAACAATGT   | 1287 |
| Db | 401  | LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys  | 420  |
| QY | 1288 | GAAGTCAGACATCAAGAACAAGAAAGGAAGATTCCAAATACATCCCACCTAACCTCCCT   | 1347 |
| Db | 421  | GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro  | 440  |
| QY | 1348 | CCAGAGAAGAGTGGCTGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGG     | 1407 |
| Db | 441  | ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly  | 460  |
| QY | 1408 | ATCCATTTCCTTAAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTT      | 1467 |
| Db | 461  | IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu  | 480  |
| QY | 1468 | CTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTTGTGGGATAAC | 1527 |
| Db | 481  | LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn  | 500  |
| QY | 1528 | AAACATGTGAAATGCCTTGTTCGGAACAAAACCTTGTACCTGTGGAAGATGAGAATGGT   | 1587 |
| Db | 501  | LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly  | 520  |
| QY | 1588 | GAGCGAACTGCAGGAGTAGGTGGGAGTCAATTCAGACTGCACCTTCTCAACAAATTCA    | 1647 |
| Db | 521  | GluArgThrAlaGlySerArgTyrGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr  | 540  |
| QY | 1648 | CGACCCCGAAGTGAAGGATCGGATTCGAAATACAAATGTGGCATATTTCTAAGAAATGG   | 1707 |
| Db | 541  | ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp  | 560  |
| QY | 1708 | GACTTTACAGCTTGTGTTGATTCTCTGGGATAAGGTACTTTGAAGAGCAGAGGCCCAACAT | 1767 |
| Db | 561  | AspPheThrAlaLeuValAspPheTyrAspLysValLeuGluGluAlaGluAlaGlnHis  | 580  |
| QY | 1768 | TTATATCAGTCCATTTTACCTGACATGGTGAATAATGCACTCTGTCTGCCAAATATTTGC  | 1827 |
| Db | 581  | LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys  | 600  |
| QY | 1828 | ACCCAGCCAATACCACTCCTGAAACAGAGATGAATCATTTCTGCAGATGTCACAGGAA    | 1887 |
| Db | 601  | ThrGlnProIleProLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu     | 620  |
| QY | 1888 | CAGATCGCCAGTCTTTTAGTAAATAGTCTTCTCTGTCACATTTCCCGCAGGAATGCCAAG  | 1947 |
| Db | 621  | GlnIleAlaSerLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys     | 640  |
| QY | 1948 | ATGAAATCGGAGTATTCTAGTTACCCAGACATTAATCTCAATCGGTGTTTGAAGGACGT   | 2007 |
| Db | 641  | MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg  | 660  |
| QY | 2008 | TCATCAAGGAAACCAAGAAACCTGAAACACTCTTCTGCTACTTTTCGAAGAGTCACAGAG  | 2067 |
| Db | 661  | SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu  | 680  |
| QY | 2068 | AAAAAACCTACAGGATTGGTGACATTTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGG | 2127 |
| Db | 681  | LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp  | 700  |

QY 2128 GAAAGTGTGAAAGCCCTCTGACACGCTTACACGCTTACACGCTTACGAGGGTACCATAGAGGC 2187  
DB 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGGCCGAGGCATGCTACAGGTGGATTGTTGCAAAATCGTTTTGTTGGAGGTGGTGACT 2247  
DB 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGACTTGTACAAGAAGAAATCAGATTGTTTAATCAATCCTGAATTGATTGTTCA 2307  
DB 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTCTACTGAGGTGCTGGATCACAATGAGTGTCTTATTATACAGGTACTGAACAG 2367  
DB 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGGCTATGCTGAACTTATCGTTGGCCCGAAGCCATGAAGATGG 2427  
DB 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTyrAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAGACGATGTCGCGGCTGCACGGAGATCGTTGTCATTGACGCACCTTAC 2487  
DB 801 SerGluLysAspAspTyrGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACCTCGATCAGTTTGTGCTGAGAAAGTGAGACGTGAGCTTAACAAGGCT 2547  
DB 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840  
QY 2548 TACTGCGGATTCTCTCGTCTGGAGTCTCTTCTGAAATCTTCTGAGTGGCCACGGGA 2607  
DB 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGCTGTGGTGCCTTTGGGGTGACGCTAGATTAAAGCCTTAATACAGATCTTG 2667  
DB 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
QY 2668 GCAGCTGCTGCGGTGAACGTGAGTGGTGTATTATTCACCTTTGGGACTCAGAGTTGATG 2727  
DB 881 AlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTTGAAAGTG 2787  
DB 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
QY 2788 TACAAGTTATTGCTTAGATACATAATGAAGATGCAGAAACTGTTCCACCCCTGGACCA 2847  
DB 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTATCCATTATATACCATGCTGTTGAGTCAAGTGCAGAGACCACTGAC 2907  
DB 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
QY 2908 ATGCCAGGACAGAGGCAGGCACC 2931  
DB 961 MetProGlyGlnLysAlaGlyThr 968  
RESULT 3  
AAU76013  
ID AAU76013 standard; protein; 968 AA.  
AC AAU76013;  
XX  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.

OS Mus musculus.  
XX US6333148-B1.  
PN  
XX  
PD 25-DEC-2001.  
XX  
PF 30-APR-1999; 99US-00302812.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-153820/20.  
DR N-PSDB; ABK14495.  
XX  
PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
PS Claim 3; Col 63-68; 80pp; English.  
XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the mouse PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX  
SQ Sequence 968 AA;  
Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 5164.00 Matches: 968  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 76.24% Indels: 0  
DB: 5 Gaps: 0  
US-09-302-812-5 (1-3814) x AAU76013 (1-968)  
QY 28 ATGAGTGGGGCCCCCGCTGGAGCCCTGCACGAAGCGCGCTGGGGCGCGTGGAACT 87  
DB 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20  
QY 88 TCTGCGCCGACTGCTCGGACTCCCGAGCTTCCCTGGCAGGCGAGCGGTGTTCTCGAC 147  
DB 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeuAsp 40  
QY 148 CCCAAGGACGCTCCCGTCCAGTTCCAGGTCCCTCCGTCCTCGCCAGCCTCGCTCGGG 207  
DB 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
QY 208 CGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTCACAAAGAACTATTACT 267  
DB 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
QY 268 ACTTGATGGATACTAAAGGACCCCAAGACAGCTGAATCAGAAAGTAAAGAAACAACAAT 327  
DB 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
QY 328 ACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTG 387  
DB 101 ThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyrProHisLysVal 120

QY 388 GAAAAATTGGAAATGTTCTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGTTCA 447  
Db 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer 140  
QY 448 CAGTATTTGAACCAACAGCAGACTGCGAGTGTGTCAAGTGGCAGAAATGAAGGAAGCAT 507  
Db 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160  
QY 508 GCAGAACAGCTTTTGGCAAGTGAGCCTCCCGGGGAGCTCCGCTACCAAGCAGCTTAGT 567  
Db 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
QY 568 AATGCTAACATTGGTCAGTCACCCACACTGATGACCAACAGTGACACAGATCATGAAGAA 627  
Db 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200  
QY 628 GACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCCAACAGTA 687  
Db 201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220  
QY 688 GGAGATGGGCAGGCCAGAAAGCAACTGTAAAGTCAGTGGATCTCGCCAGTCTGTGAAAGAC 747  
Db 221 GlyAspGlyGlnAlaArgSerAsnCysLysSerGlySerArgGlnSerValLysAsp 240  
QY 748 TGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACACAGAGATCCTTTGTGATGTT 807  
Db 241 CysThrGlyCysGlnGlnGluGluValAspValLeuProGluSerProLeuSerAspVal 260  
QY 808 GGTCCGAGGACATTTGGAAGTGGACCAAAAATGACAAACAATTTGACTGGACAAAGAAC 867  
Db 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280  
QY 868 AGCCTAGGTGATCGCCTCCATTGAGAAAGAAAGTGAGCCTGAGTCAACCAATGGATGTA 927  
Db 281 SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal 300  
QY 928 GACAACCTCGAAGAACAGTTGTCAAGATTCAAGATTCAGAACAGATGAAGAAACAAGTCCAGTCTTT 987  
Db 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
QY 988 GATGAGCAAGATGATCGTTCTCCTCCCAACAGCAAAATAAACTTTCAAGTTGCCAAGCAAGA 1047  
Db 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340  
QY 1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107  
Db 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
QY 1108 TTCCAAATTTGAGGAGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGA 1167  
Db 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
QY 1168 AACTCTTTAGCCTTAATGTAGAGTGTAGAGTTCCAAGCAGCATGGAAAAAGGATTCT 1227  
Db 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
QY 1228 AAAATTACAGATCATTTTCATGAGAAATTTCCAAGTCAGAGGACAGAGAAAGAAACAATGT 1287  
Db 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420  
QY 1288 GAAGTCAGACATCAAGAAACAGAAAGGAGATTCCTTAATAATACATCCACCTAACCTCCCT 1347  
Db 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440  
QY 1348 CCAGAGAGAAAGTGGTGGGAACCTCCTATTGAGGAAATGAGAAAAATGCGTGTGGG 1407  
Db 441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460  
QY 1408 ATCCATTTGCCCTTCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCCGGTAGACCTT 1467  
Db 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
QY 1468 CTGAGAGCAGGAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTTGTGGGATAAC 1527

Db 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500  
QY 1528 AAACATGTGAAAAATGCTTGTTCGGAAACAAACTTGTACCCCTGTGGAGATGAGAAATGGT 1587  
Db 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
QY 1588 GAGCGAACTGCAGGGAGTAGTGGGAGCTCATTCAGACTGCATCTCTCAACAAATTCACA 1647  
Db 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540  
QY 1648 CGACCCCAAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCAATTCCTAAGAAATGG 1707  
Db 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560  
QY 1708 GACTTTACAGCTTTGGTTGATTCTGGGATAAGGTACTTTGAAGAGCAGAGGCCCAACAT 1767  
Db 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGlnHis 580  
QY 1768 TTATATCAGTCCATTTTACCTGACATGCTGTGAAAAATTCACACTCTGTCTGCCAAATATTTGC 1827  
Db 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCGAGCCAATACCACCTCCTGAAACAGAAAGATGAATCATTTCTGTACGATGTCACAGGAA 1887  
Db 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATCGCCAGTCTTTTAGTAAATGCTTTCTTCTGCACATTTCCCGACGGAATGCCAAG 1947  
Db 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640  
QY 1948 ATGAAATCGGAGTATTCTAGTTACCCAGACATTAACCTTCAATCGGTTGTTTGAAGGACGT 2007  
Db 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAAGGAAACCCAGAAAACTGAAAAACACTTCTGTCTACTTTCCGAAGAGTCACAGAG 2067  
Db 661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu 680  
QY 2068 AAAAACTACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGG 2127  
Db 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp 700  
QY 2128 GAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACAGAGGTACCATAGAAGGC 2187  
Db 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGGCCGAGGCATGCTACAGTGGATTTTGCAAATCGTTTTTGTGGAGGTGGTGTGACT 2247  
Db 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGACTTGTACAAGAAAGAAATCAGATTTTAAATCAATCAATCTGAATTTGTTTCA 2307  
Db 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATATATCAGAGGTACTGAACAG 2367  
Db 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCCGAAGCCATGAAGATGGG 2427  
Db 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAGGACGATTGGCAGCGCGCTGCACGGAGATCGTTGCCATTGACGCACCTTCAC 2487  
Db 801 SerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACCTCGATCAGTTGTGCTGCTGAGAAAGTGAGACGTGAGCTTAACAAGGCT 2547  
Db 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840  
QY 2548 TACTCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAATCTTCTGAGTGGCCACGGGA 2607



Db 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGGCTGTGGTGCCTTTGGGGTGGCGTACGCTAGATTAAAGCCTTAATACAGATCCTG 2667  
Db 861 AsnTrpGlyCysGlyAlaPheGlyGlyGlyAlaArgLeuLysAlaLeuIleGlnLeu 880  
QY 2668 GCAGCTGTGGCTGAACGTGACGTGGTGGTTATTTCACCTTTGGGGACTCAGAGTTGATG 2727  
Db 881 AlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGACACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTTGGAAAAGTG 2787  
Db 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
QY 2788 TACAAGTTATTGCTTAGATACATCAATGAAGATGCAGAAACTGTTCCACCCCTGGACCA 2847  
Db 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTATCCATTATATACCATGCTGTTGAGTCAAGTGCAGAGACCACCTGAC 2907  
Db 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
QY 2908 ATGCCAGACAGAAAGCGAGGCACC 2931  
Db 961 MetProGlyGlnLysAlaGlyThr 968

RESULT 4

ABG72280  
ID ABG72280 standard; protein; 968 AA.

XX AC ABG72280;

XX 13-MAR-2003 (first entry)

DE Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.

XX Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
KW cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiant;  
KW vasotropic; anticonvulsant; cerebroprotective; enzyme.

XX Mus musculus.

XX US2002132328-A1.

PD 19-SEP-2002.

XX 09-OCT-2001; 2001US-00973451.

PR 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

XX (JACO/) JACOBSON M K.

PA (JACO/) JACOBSON E L.

PA (AMEJ/) AME J.

PA (LINW/) LIN W.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

DR WPI; 2003-155895/15.

DR N-PSDB; ABX14479.

XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.

XX Claim 28; Fig 16; 86pp; English.

PS

CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents murine PARG enzyme

XX SQ Sequence 968 AA;

Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 5164.00 Matches: 968  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 76.24% Indels: 0  
DB: 6 Gaps: 0

US-09-302-812-5 (1-3814) x ABG72280 (1-968)

QY 28 ATGAGTGGGGCCCCCGCTGGAGCCCTGCACGAAAGCGCGTGGGGCCCGCTGGAAC 87  
Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLysAlaArgTyrGlyAlaAlaGlyThr 20

QY 88 TCTGCGCGGACTGCTCGGACTCCCGGAGCTTCCCTGGCAGGAGCGGTCTCGAC 147  
Db 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeuAsp 40

QY 148 CCCAAGGACGCTCCGTCAGTTCAGGTCCCTCCGTCCCTCGCCAGCGCTCGTTCGGGG 207  
Db 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60

QY 208 CGGGCGGACCGCACAGAGGCAACGCCACCTCGTTGTTTTCAAAACAAAAGACTATTACT 267  
Db 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80

QY 268 ACTTGATGATATACTAAAGGACCCCAAGCAGCTGAATCAGAAAGTAAAGAAAACAACAAT 327  
Db 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100

QY 328 ACAAGAATTGACTCCATGATGAGTTCCTGTCAGAAAGATAACTTTTCCCATAGAGTG 387  
Db 101 ThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyrProHisLysVal 120

QY 388 GAAAAATTGAAAAATGTTCTCAGCTAAATCTGATAAATCACCACAGAAAAGAGTTCA 447  
Db 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer 140

QY 448 CAGTATTTGAAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGATGAAGGAGCAT 507  
Db 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160

QY 508 GCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAGCAGCTTAGT 567  
Db 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180

QY 568 AATGCTAACATTGGTCAGTCACTCCACACTGATGACACAGTGACACAGATCATGAAGAA 627  
Db 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200

QY 628 GACAGAGACAATCAGCAAGTTCTTACACCTATAAACTGCAATACAAAGCCACAGTA 687  
Db 201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220



QY 688 GGAGATGGCAGGCCAGAAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGAC 747  
Dbb 221 GlyAspGlyGlnAlaArgSerAsnGlyCysSerGlySerArgGlnSerValLysAsp 240  
QY 748 TGTACAGGCTGTCAACAGGAGGAGGTGATGTCTACAGAGAGTCTTTGTGATGTT 807  
Dbb 241 CysThrGlyCysGlnGlnGluGluValAspValLeuProGluSerProLeuSerAspVal 260  
QY 808 GGTGCCGAGGACATTGGAACTGGACCCAAAAAATGACAAATAATGACTGGACAAGAAC 867  
Dbb 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280  
QY 868 AGCCTAGGTGATTCGCCCTCCATTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTA 927  
Dbb 281 SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal 300  
QY 928 GACAACTCGAGAAACAGTTGTCAAGATTGAGAACGAGATGAAGAAACAAGTCCAGTCTTT 987  
Dbb 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
QY 988 GATGAGCAAGATGATCGTTCTCCCAACAGCAAAATAAATTTCAAGTGCACCAAGCAAGA 1047  
Dbb 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340  
QY 1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107  
Dbb 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
QY 1108 TTCCAATTTGAAGGAGAAAAATAATGCTGGGACCACTGACTTAATGCCAAGCCATCTGGA 1167  
Dbb 361 PheGlnPheGluGlyGluAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAGTTCCAAGCAGCATGGAAGGATCTCT 1227  
Dbb 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
QY 1228 AAAATTACAGATCAATTTCATGAGAAATTTCCAAGTCAGAGGACAGAAAGAAACAATGT 1287  
Dbb 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420  
QY 1288 GAAGTCAGACATCAAGAAACAGAAAGGAAGATTCAGAAATACATCCACCTAACCTCCCT 1347  
Dbb 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440  
QY 1348 CCAGAGAAGAAGTGGTGGAACTCCTATTGAGGAATGAGAAATGCCTCGGTGGG 1407  
Dbb 441 ProGluLysLysTyrLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460  
QY 1408 ATCCATTTGCCCTTCCCTTAAGACCATCTGCAAGTCACAGTGACTGTTTCGGGTAGACCTT 1467  
Dbb 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
QY 1468 CTGAGAGCAGGAGAGGTTCGAAACCTTTTCCAACACATTAACAAGATTGTGGGATAAC 1527  
Dbb 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTyrAspAsn 500  
QY 1528 AAACATGTGAAAATGCCCTTGTTCGGAACAAACCTTGACCTGTGGAGATGAGAAATGGT 1587  
Dbb 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
QY 1588 GAGCGAACTGCAGGAGTAGTGGGAGCTCATTCAGACTGACCTTCTCAACAATTCACA 1647  
Dbb 521 GluArgThrAlaGlySerArgTyrGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540  
QY 1648 CGACCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAATGG 1707  
Dbb 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560  
QY 1708 GACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGACAGAGGCCCAACAT 1767  
Dbb 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGlnHis 580  
QY 1768 TTATATAGTCCATTTTACCTGACATGGTGAAATTTGCACCTCTGTCTGCCAAATATTTGC 1827

Dbb 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCAGCCAAATACCACTCCTGAAACAGAAAGATGAATCATCTGTGTACGATGTACAGGAA 1887  
Dbb 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCGACGGAATGCCAAG 1947  
Dbb 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgAsnAlaLys 640  
QY 1948 ATGAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTGTTTGAAGGACGT 2007  
Dbb 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAAGGAAACAGAAATACTGAAACACACTCTTCTGCTACTTTCGAAGAGTCACAGAG 2067  
Dbb 661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu 680  
QY 2068 AAAAACTACAGGATTGGTGACATTACAAGACAGAGTCTTGAAGATTTTCCAGAATGG 2127  
Dbb 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp 700  
QY 2128 GAAAGGTCTGAAAAGCCTCTGACACGCTTACACGTCACCTTACGAGGGTACCATAGAAGGC 2187  
Dbb 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGGCCGAGGCATGCTACAGGTGGATTTTGCAATCGTTTGTGGAGGTGGTGTGACT 2247  
Dbb 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGGACTTGTACAAGAAGAAATCAGATTTTAAATCAATCCTGAATGATTGTTTCA 2307  
Dbb 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACAG 2367  
Dbb 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGCCCCGAAGCCATGAAGATGGG 2427  
Dbb 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTyrAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAGAAGACGATTTGGCAGCGGCTGACGAGATCGTTGCCATTTGACGCTTCCAC 2487  
Dbb 801 SerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACCTCGATCAGTTTGTGCTGAGAAAGTGAGACGTGAGCTTAACAAGCT 2547  
Dbb 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840  
QY 2548 TACTGCGGATTCCTCCGTCTGGAGTTCCTTCTGAAAAATCTTTCTGAGTGGCCACGGGA 2607  
Dbb 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGCTGTGCTGCTTTGGGGGTGACGCTAGATTAAGAAGCCTTAATACAGATCCTG 2667  
Dbb 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
QY 2668 GCAGCTGCTGGCGTGAACGTGACGTGGTTTATTTACCTTTGGGGACTCAGAGTTGATG 2727  
Dbb 881 AlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGAGACATTTACAGATGCACACTTTTCTTACCGAGAGGAAGCTGGATGTTGGAAAAGTG 2787  
Dbb 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
QY 2788 TACAAGTTATTGCTTAGATACATAACAATGAAGATGAGAAACTGTTCCACCTCGGACCA 2847  
Dbb 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTTATCCATTATATACCATGCTGTTGAGTCAAGTGCAGAGACCACTGAC 2907



QY 1084 AAGGGAAGTCAAGTTAGATTGCAATTTCCAAATTTGAA---GGAGAAATAATAGTCTGGGACC 1140  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAAGCAGCATGGAAAAAGGATTTCTAAATACAGATCATTTCTAGAAATTTCCAAG 1260  
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAGAAAAAGCAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATT 1320  
Db 420 AlaGluAspArgArgLysGluGlnTyrGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1321 CCAAAATACATCCACCTAACCTCCCTCCAGAGAGAAAGTGGCTGGGAACCTCTATTGAG 1380  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTyrLeuGlyThrProIleGlu 459  
QY 1381 GAAATGAGAAAAATGCTCGGTGGGTGATCCATTTGCCTTCCTTAAGACCATCTGCAAGT 1440  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACACAGTGACTGTTGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGGAACCTTTTCCA 1500  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1501 ACACATTACAAAGATTTCGGGATAACAAACATGTGAAATGCCTTGTTCCGGAACAAAC 1560  
Db 500 ThrHisTyrLysAspLeuTyrAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1561 TTGTACCTGTGGAAGATGAGAAATGGTGAGCGAACTGCAGGGAGTAGTGGGAGCTCATT 1620  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTyrGluLeuIle 539  
QY 1621 CAGACTGCATCTCTCAACAAATTTCACAGCCCCAGAACTTGAAGGATGCGATTCTGAAA 1680  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740  
Db 560 TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuIleAspPheTyrAspLys 579  
QY 1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAA 1800  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCACTCTGTCTGCCAAATATTTGACCCAGCCCAATACACTCCTGAAACAGAGATG 1860  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATCTGTCCAGATGTCAGGAATGCCAAGATGAAATCGGAGTATTTAGTACCCAGACATT 1920  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TGCACATTTCCCGACGGAATGCCAAGATGAAATCGGAGTATTTAGTACCCAGACATT 1980  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspile 659  
QY 1981 AACTTCAATCGGTTGTTGAAGGACGTTTCATCAAGGAAACCCAGAAAACTGAAACACATC 2040  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAACCTACAGGATTGCTGACATTTTACAAGA 2100  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2101 CAGAGTCTTGAGATTTTCCAGAATGGAAAGGTGTGAAAGCCTCTGACACGCTTACAC 2160  
Db 700 GlnSerLeuGluAspPheProGluTyrGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2161 GTCACCTACGAGGGTACCATAGAGGCAACGGCCGAGGCATGCTACAGGTGATTTTGCA 2220

Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2221 AATCGTTTGTGGAGGTGGTGTGACTGGTCCGGGACTTGTACAGAAAGAAATCAGATT 2280  
Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759  
QY 2281 TTAATCAATCCTGAATTGATTGTTTACGGCTGTTCACTGAGGTGCTGGATCACAATGAG 2340  
Db 760 LeuIleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2401 CGTTGGCCCGAAGCCATGAAGATGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACG 2460  
Db 800 ArgTyrSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2461 GAGATCGTTGCCATTGACGCACCTTCACTTCAGACGCTACCTTCGATCAGTTTGTGCTGAG 2520  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 839  
QY 2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCTCCGTCCTGGAGTTCCTTCT 2580  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2581 GAAAATCTTCTGAGTGGCCACGGGAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2641 AGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGTGAACGTCGCTGTTTAT 2700  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTCTTACC 2760  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 2761 GAGAGGAAGCTGGATGTTGAAAAGTGTACAGTATTGCTTAGATACATACTACAATGAAGAA 2820  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTTCATATACCATGCT 2880  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 2881 GTTGAGTCAAGTGCAGACCACTGACATGCCAGACAGAGGACGAGCCACC 2931  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 6

AAU76021

ID AAU76021 standard; protein; 976 AA.

XX AC AAU76021;

XX DT 08-MAY-2002 (first entry)

XX DE Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.

XX KW Human; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.

OS Homo sapiens.

XX US6337202-B1.

XX PD 08-JAN-2002.



XX PF 23-FEB-2000; 2000US-005111477.  
XX PR 01-MAY-1998; 98US-0083768P.  
XX PR 30-APR-1999; 99US-00302812.  
XX PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX DR WPI; 2002-163240/21.  
XX DR N-PSDB; ABK14932.  
XX PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX PS Claim 2; Col 55-60; 81pp; English.  
XX CC The present invention relates to a new poly (ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the human PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX CC  
SQ Sequence 976 AA;

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 64.89% Indels: 10  
DB: 5 Gaps: 6

US-09-302-812-5 (1-3814) x AAU76021 (1-976)

QY 28 ATGAGTGGGGCCCGCTGGGAGCCCTGCACGAAGCG---CGCTGGGGCGCGCTGGA 84  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19  
QY 85 ACTTCTGCGCGGACTGCTCGGACTCCCGGAGCTTCCCTGGCAGGCGGCTGCTC 144  
Db 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGCTCCGCGAGCTCGCTCG 204  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGGCGGACCGGCACAGAGGCAACGCCACCTCGTTTCTTTCAACAAAGACTATT 264  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 265 ACTACTGGATGGATACTAAAGGACCCCAAGACAGCTGAATCAGAA-----AGTAAA 315  
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 316 GAAACAAACAATAACAAGATTGACTCCATGATGAGTTCTGTGAGAAAGATACTTTTAC 375  
Db 100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
QY 376 CCACATAAGTGGAAAAATTGGAAAAATGTTCTCCTCAGCTAAATCTTGATAAAATCACCCACA 435

Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 436 GAAAGAGATTCAAGTATTGTAACCAACACAGCAGACTCGGAGTGTGTGCAAGTGCAGAAAT 495  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 496 GAAGGGAAGCATGCAGAAACAGCTTTTGGCAAGTGAGCCTCCCGGGGACTCCGCTACCA 555  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 556 AAGCAGCTTACTAATGCTAATGCTAATGCTAATGCTAATGCTAATGCTAATGCTAATGCTA 615  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 616 GATCATGAAGAGACAGACAGCAATCAGCAGTTTCTTACACCTATAAACTTGCATAATACA 675  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 676 AAGCCAAACAGTAGAGATGGGCGGCGCAGA-----AGCAACTGTAACTGTCAGTGA 726  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGCCAGTCTGTGAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCA 786  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
QY 787 GAGAGTCTTTGTACAGATGTTGGTCCCGAGGACATTTGGAACCTGGACCAAAAATGACAAC 846  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 847 AAATTGACTGGACAAGAAAGCAGCCTAGTGATTCGCTCCCTCCCTCCCAACAGCAAAAT 906  
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299  
QY 907 CCTGAGTCACCAATGGATGTAGACAACCTCGAGAAACAGATTGTCAAGATTTCAGAAAGCAG 966  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAAAT 1023  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
QY 1024 AAATTCTCAAGTTGCCAAGCAAGCAAGCTGATGCGGATCTTAGGAAACCGTATTGACT 1083  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1084 AAGGGAAGTGAAGTTAGATTGATTCATTTCCAAATTTGAA---GGAGAAAATAATGCTGGGACC 1140  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAAATGCCAAGCCCATCTGGAACCTCTTACGCTTAATGTAGAGTGAGAGT 1200  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAAGCAGCATGGAAAAGGGATTCTTAAATTTACAGATCATTTTCATGAGAAATTTCCAAG 1260  
Db 400 SerLysGlnHisGlyLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAAAGAAACAATGTGAAGTCAGACATCAAGAACAGAAAGGAAAGATT 1320  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1321 CCAAAATACATCCACCTAACCTCCCTCCAGAGAGAGAGTGGCTGGGAACCTCCTATTGAG 1380  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1381 GAAATGAGAAAAATGCCTCGGTGGGATCCATTTGCTTCTTAAAGACCATCTGCAAGT 1440  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACACAGTGACTGTTTCGGGTAGACCTTCTTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProphePro 499



QY 1501 ACACATTACAAAGATTCTGGGATAACAAACATGTGAAAATGCTTGTTCGGAACAAAC 1560  
Db 500 ThrHisTyrLysAspLeuTyrPaspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1561 TTGTACCTGTGGAAGATGAGAAATGGTGAGCGAACTGCGAGGGAGTAGGTGGAGCTCAT 1620  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTyrPaspGluLeu 539  
QY 1621 CAGACTGCATCTCTCAACAAATTCACAGACCCAGAACTTGAGAGATGCGATTCTGAAA 1680  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaLeuLys 559  
QY 1681 TACAATGTGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGATAAG 1740  
Db 560 TyrAsnValAlaTyrSerLysLysTyrPaspPheThrAlaLeuLeuAspPheTyrPaspLys 579  
QY 1741 GTACTTGAAGAGCAGAGCCCAACATTTATATATATATATATATATATATATATATATAT 1800  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCACCTGTCTGCCAAATATTTGCACCCAGCCCAATACCACTCTCTGAAACAGAGATG 1860  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATTTCTGCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGTAAATGCTTTCTTC 1920  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TGCACATTTCCCGACGGAATGCCAAGATGGAATCGGAGTATTTAGTACCCAGACATT 1980  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 1981 AACTTCAATCGGTGTTTGAAGGACGTTTCATCAAGGAAACCCAGAAAACCTGAAAACACTC 2040  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 679  
QY 2041 TTCTGTACTTTCGAAGAGTCACAGAGAAAACCTACAGGATTGGTGACATTTACAAGA 2100  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2101 CAGAGTCTTGAAGATTTTCCAGAAATGGGAAGGTGTGAAAAGCCTCTGACACGCTTACAC 2160  
Db 700 GlnSerLeuGluAspPheProGluTyrPaspGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2161 GTCACATTACAGGGTACCATAGAAGGCAACGGCGAGGCATGCTACAGGTGGATTGCA 2220  
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2221 AATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGACTTGTACAAGAAAGAAATCAGATT 2280  
Db 740 AsnArgPheValGlyGlyValThrSerArgLeuPheThrGluValGlnGluArgPhe 759  
QY 2281 TTAATCAATCCTGAATGATTTGTTTACGGCTGTTTCACTGAGGTGCTGGATCACAATGAG 2340  
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2341 TGTCTTATTATCAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2401 CGTTGGGCGGAGCCATGAAGATGGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACG 2460  
Db 800 ArgTyrSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2461 GAGATCGTTGCCATTGACGCACTTCACTTACAGCGCTACCTCGATCAGTTTGTGCTGAG 2520  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2521 AAAGTGAGACGTGAGCTTAAACAAGCTTACTGCGGATTCCTCCGTCTGGAGTTCCTTCT 2580  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

2581 GAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGCTGCTTGGGGGTGACGCT 2640  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2641 AGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGGCGCTGAACGTGACGTGGTTTAT 2700  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 899  
QY 2701 TTCACTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 2761 GAGAGGAAGCTGGATGTTGGAAGTGTACAAGTTATTGCTTAGATACATAATGAAGAA 2820  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 2821 TGCAGAAAACCTGTTCCACCCCTGGACACAGACATCAAGCTTTATCCATTTCATATACCATGCT 2880  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAGGAGGAGGCACCC 2931  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976  
RESULT 7  
AAU76012  
ID AAU76012 standard; protein; 976 AA.  
XX AC AAU76012;  
XX 08-MAY-2002 (first entry)  
DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX Homo sapiens.  
XX US6333148-B1.  
XX 25-DEC-2001.  
XX 30-APR-1999; 99US-00302812.  
XX 01-MAY-1998; 98US-0083768P.  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-153820/20.  
XX N-PSDB; ABK14494.  
XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
XX useful potentially for treating diseases associated with DNA damage, e.g.  
XX cancer.  
XX Claim 3; Col 55-60; 80pp; English.  
XX The present invention relates to a new method for screening compounds for  
XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
XX diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
XX inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
XX are used to treat or prevent any condition associated with DNA damage  
XX e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
XX vascular stroke, aging and neurodegeneration e.g. Huntington's,  
XX Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
XX Compounds identified by the new method are more effective than known  
XX inhibitors and have fewer side effects. The present amino acid sequence

CC represents the human PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX Sequence 976 AA;  
SQ Sequence 976 AA;  
Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 64.89% Indels: 10  
DB: 5 Gaps: 6  
US-09-302-812-5 (1-3814) x AAU76012 (1-976)  
QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCACGAAGCG---CGTGGGGCGCGCTGGA 84  
DB 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19  
QY 85 ACTTCTGCGCGACTGCTCGGACTCCCGGAGCTTCCTGGCAGGCAGAGCGCTGTTCTC 144  
DB 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 145 GACCCCAAGGACGCTCCCGTCCAGTTTCAGGCTCCCTCCGTCCTCGCCAGCCCTGCTCG 204  
DB 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGGGGACCGGACAGAGGCAACGCCACCTCGTTGTTTCAAACAAGACTATT 264  
DB 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 265 ACTACTTGATGGAVACTAAAGGACCCAGACAGCTGAATCAGAA-----AGTAAA 315  
DB 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 316 GAAACAACAATACAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTAC 375  
DB 100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
QY 376 CCACATAAGCTGGAATAATTGGAATAATGTTCTCCTCAGCTAAATCTTGATAAATCACCACA 435  
DB 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 436 GAAAAGAGTTCACAGATTTGAACCAACAGCAGACTCGGAGTGTGTGCAAGTGGCAGAAT 495  
DB 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 496 GAAGGGAAGCATGCAAGACAGCTTTTGGCAAGTGAAGCTCCCGGGGAGCTCCGCTACCA 555  
DB 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 556 AAGCAGCTTAGTAATGCTAAACATTGGTCAGTCACCCACACTGATGACCCACAGTGACACA 615  
DB 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 616 GATCATGAAGAAGACAGACACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAAATACA 675  
DB 200 AspSerGluGluAsnArgAspAsnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 676 AAGCCAACAGTAGGAGATGGGCAGGCCAGA-----AGCAACTGTAAGTGCAGTGA 726  
DB 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGCCAGTCTGTGAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTCTACCA 786  
DB 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro 259  
QY 787 GAGAGTCTTTGTGATGTTGGTGGCAGGACATTGGAACCTGGACCAAAAATGACAAC 846  
DB 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 847 AAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCTCCATTTTGAGAAAGAAAGTGAG 906

DB 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
QY 907 CCTGAGTCACCAATGGATGTAGACAACCTCAGAAACAGTTGTCAAGATTTCAGAAGCAGAT 966  
DB 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAAAT 1023  
DB 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
QY 1024 AAACCTTTCAAGTTGCCAAGCAAGAGAGCTGATGGCGATCTTAGGAAACGGTATTGACT 1083  
DB 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1084 AAGGGAAGTGAAGTTAGATTGATTTCATTTCCATTTGAA---GGAGAAAATAATGCTGGGACC 1140  
DB 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAATGCCAAGCCCATCTGGAAACTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200  
DB 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAAGCAGCATGCAAAAAGGATTCTAAATACAGATCATTTTCATGAGAAATTTCCAAG 1260  
DB 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAAAGAAACAATGTGAAGTCAGACATCAAAAGAACAGAAAGAAAGATT 1320  
DB 420 AlaGluAspArgArgLysGluGlnTrpGlnThrLysHisGlnArgThrGluArgLysIle 439  
QY 1321 CCAAAATACATCCACCTAACCTCCCTCCAGAGAAGAGTGGTGGGAACCTCCTATTGAG 1380  
DB 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1381 GAAATGAGAAAATGCCCTCGGTGGGATCCATTTCCTTAAGACCATCTGCAAGT 1440  
DB 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACACAGTGAAGTTCGGGTAGACCTTCAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500  
DB 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1501 ACACATTACAAAGATTGTGGGATAACAAACATGTGAAATGCCTTGTTCGGAAACAAAC 1560  
DB 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1561 TTGTACCCCTGTGGAAGATGAGAAATGGGACTTTTACAGCTTTGGTTGATTTCTGGGATAAG 1620  
DB 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1621 CAGACTGCACCTTCTCAACAAATTCACACGACCCCAAGAACTTGAAGGATGCGATTCTGAAA 1680  
DB 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740  
DB 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1741 GTACTTGAAGAACGAGAGGCCCAACATTATATATATATATATATATATATATATATATAT 1800  
DB 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCACCTCTGTGTCGCAAAATATTGACCCAGCCCAATACCACTCCTGTAACAGAGATG 1860  
DB 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATCTCTGTACGATGTACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTCTTCTTC 1920  
DB 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TGCACATTTCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 1980

Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659

QY 1981 AACTTCAATCGGTTGTTGAAGGAGCGTTTCATCAAGGAAACCAAGAAAACTGAAAAACACTC 2040

Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679

QY 2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAACCTACAGGATTGGTGACATTTCACAAGA 2100

Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699

QY 2101 CAGAGTCTTGAAGATTTCAGAAATGGGAAAGGTGTGAAAAGCCTCTGTGACACGCTTACAC 2160

Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719

QY 2161 GTCACCTACGAGGTTACCATAGAGGCAACGCGCGAGGATGCTACAGGTGGATTGCA 2220

Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739

QY 2221 AATCGTTTGTGGAGGTGGTGTGACTGCTGCGGACTTGTACAGAAGAAATCAGATT 2280

Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759

QY 2281 TTAATCAATCCTGAATTGATTGTTTTCACGCTGTTCTCACTGAGGTGCTGGATCACAATGAG 2340

Db 760 LeuIleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400

Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799

QY 2401 CGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACG 2460

Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819

QY 2461 GAGATCGTTGCAATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCTGAG 2520

Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

QY 2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTCGGATTCTCCTCCGTCCTGGAGTTCTTCT 2580

Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

QY 2581 GAAATCTTTCTGCAGTGGCCACGGGAACTGGGCTGTGGTGCCTTGGGGGTGACGCT 2640

Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879

QY 2641 AGATTAAAGCCTTAATACAGATCCTGGAGCTGCTGCGGTGAACGTGACGTGTTTAT 2700

Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899

QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTACAGCATGACACTTTCCTTACC 2760

Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

QY 2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACATAAGAA 2820

Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuArgTyrTyrAsnGluGlu 939

QY 2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTTCATATACCATGCT 2880

Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

QY 2881 GTTGAGTCAAGTGAGAGACCACTGACATGCCAGGACAGAAGCGGACCC 2931

Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 8

ABG72279

ID ABG72279 standard; protein; 976 AA.

XX

AC ABG72279;

XX

DT 13-MAR-2003 (first entry)

XX DE Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.

XX KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;

KW cellular response; DNA damage; neoplastic disorder inducing agent;

KW oxidative stress; neoplastic disorder; myocardial infarction;

KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;

KW Parkinson's disease; Huntington's disease; inborn genetic error;

KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;

KW neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;

KW anticonvulsant; cerebroprotective; enzyme.

XX OS Homo sapiens.

XX PN US2002132328-A1.

XX PD 19-SEP-2002.

XX PF 09-OCT-2001; 2001US-00973451.

XX PR 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

XX (JACO/) JACOBSON M K.

PA (JACO/) JACOBSON E L.

PA (AMEJ/) AME J.

PA (LINW/) LIN W.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2003-155895/15.

DR N-PSDB; ABX14478.

XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)

PT glycohydrolase activity, for preventing, treating, or ameliorating a

PT disease condition, e.g. neoplastic disorder, myocardial infarction or

PT vascular stroke.

XX Claim 28; Fig 16; 86pp; English.

PS The present invention relates to the isolation of poly(ADP-ribose)

XX glycohydrolase (PARG) from several species, and the polynucleotide

CC sequences encoding them. Methods for inhibiting PARG expression or

CC overexpressing PARG are also disclosed. PARG is involved in the cellular

CC response to DNA damage, and is associated with the body's response to

CC neoplastic disorder inducing agents and oxidative stress. The

CC polynucleotide sequences encoding PARG and PARG modulators are useful for

CC preventing, treating, or ameliorating diseases such as neoplastic

CC disorders, myocardial infarction, vascular stroke, neurodegenerative

CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's

CC disease), inborn genetic errors, reperfusion following ischaemia, aging,

CC and neurotoxicity. The polynucleotide sequences are also useful in gene

CC therapy. The methods are useful for identifying an agent that modulates

CC PARG activity, identifying a mutant PARG allele in an individual, or

CC screening candidate molecules for PARG modulating activity. The present

CC sequence represents human PARG enzyme

XX Sequence 976 AA;

SQ

Alignment Scores:

Pred. No.: 0 Length: 976

Score: 4395.00 Matches: 837

Percent Similarity: 90.99% Conservative: 52

Best Local Similarity: 85.67% Mismatches: 78

Query Match: 64.89% Indels: 10

DB: 6 Gaps: 6

US-09-302-812-5 (1-3814) x ABG72279 (1-976)

QY 28 ATGAGTGGGGCCCCCGGCTGGGAGCCCTGCACGAAAGCG---CGTGGGGCGCCGCTGGA 84

Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19



QY 85 ACTTCTGCGCCGACTGCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCTGTCTC 144  
 Db 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
 QY 145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGCTCCCTCGCTCCGCGAGCCCTGCTCG 204  
 Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
 QY 205 GGGCGGGCGGACCGGACAGAGGCAACGCCACCTCGTTTGTGTTTCAACAAAGACTATT 264  
 Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrile 79  
 QY 265 ACTACTTGATGGATGATACTAAAGGACCCAGACAGCTGAATCAGAA-----AGTAAA 315  
 Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
 QY 316 GAAACAACAATACAGAATTGACTCCATGATGAGTCTGTGTCAGAAAGATAACTTTTAC 375  
 Db 100 GluAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
 QY 376 CCACATAAGGTGGAATAATTGAAATGTTCTCAGCTAAATCTTGATAAATCACCACACA 435  
 Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
 QY 436 GAAAAGAGTTTCAAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495  
 Db 140 GluLysSerThrGlnFyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
 QY 496 GAAGGGAAGCATGCAAGACAGCTTTTGGCAAGTGAGCCTCCCGGGGAGTCCGCTACCA 555  
 Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
 QY 556 AAGCAGCTTAGTAATGCTAAACATTGTCAGTCAACCCACACTGATGACCCAGTGCACACA 615  
 Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
 QY 616 GATCATGAAGAAGACAGACACAATCAGCAGTTCCTTACACCTATAAACTGCAATACA 675  
 Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
 QY 676 AAGCCCAACAGTAGAGATGGCGAGGCCAG-----AGCAACTGTAGTGCAGTGA 726  
 Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
 QY 727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTCTACCA 786  
 Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
 QY 787 GAGAGTCTTTGTGATGTTGGTCCGAGGACATTGGAACCTGGACCCAAATAATGACAAC 846  
 Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
 QY 847 AAATTGACTGGACAAGAAAGCAGCCTAGTGTATTCGCTCCATTTGAGAAAGAAAGTGAG 906  
 Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
 QY 907 CCTGAGTCACCAATGGATGTAGACAACCTCGAGAAACAGTTGTCAAGATTTCAGAACAGAT 966  
 Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
 QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCCTCCCAACAGCAAAT 1023  
 Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
 QY 1024 AAACCTTTCAAGTTGCAAGCAAGAGAGCTGATGCGCATCTTAGGAAACGGTATTTGACT 1083  
 Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
 QY 1084 AAGGGAAGTGAAGTTAGATTGCAATTTCCAAATTGAA---GGAGAAATAATGCTGGGACC 1140  
 Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
 QY 1141 AGTGACTTAAATGCCAAGCCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200

Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
 QY 1201 TCAAGCAGCATGGAAGAAAGGATTCTAAATATACAGATCATTTTCATGAGAAATTTCCAAG 1260  
 Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
 QY 1261 TCAGAGGACAGAAAGAAAGCAATGTGAAGTCAGACATCAAGAAACAGAAAGGAAGATT 1320  
 Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
 QY 1321 CCAAAATACATCCACCTAACCTCCCTCCAGAGAAAGTGGCTGGGAACCTCTATTGAG 1380  
 Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 459  
 QY 1381 GAAATGAGAAATGCTCGGTGGGTGATCCATTTGCCTTCCCTTAAGACCATCTGCAAGT 1440  
 Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
 QY 1441 CACACAGTGAATGCTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTCCA 1500  
 Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
 QY 1501 ACACATTACAAAGATTGTGGGATAACAAACATGTGAAATGCCTTGTTCGGAACAAAAC 1560  
 Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
 QY 1561 TTGTACCTCTGTGGAAGATGAGAATGGTGTGAGCAACTGCAGGAGTAGTGGGAGCTCAT 1620  
 Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
 QY 1621 CAGACTGCATCTCTCAACAAATTCACACGACCCAGAACTTGAAGGATGCGATTCGAAA 1680  
 Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
 QY 1681 TACAATGTGCATATTCTAAGAAATGGGACTTTTACAGCTTTGGTTGATTCTGGGATAAG 1740  
 Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
 QY 1741 GTACTTTGAAGACAGAGGCCCAACATTTATATATATATATATATATATATATATATAT 1800  
 Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
 QY 1801 ATTGCACCTCTGTCTGCCAAATATTTTGACCCAGCAATACCACTCTCTGAAACAGAAAGATG 1860  
 Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
 QY 1861 AATCATTTCTGTACAGATGTACAGGAACAGATCGCCAGTCTTTTAGCTAATCTTCTTTC 1920  
 Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
 QY 1921 TGCACATTTCCCGACGGAATGCCAAGATGCAATCGGAGTATTTCTAGTTACCCAGACATT 1980  
 Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerSerGluTyrSerSerTyrProAspIle 659  
 QY 1981 AACTTCAATCGGTTGTTGAAGGACGTTTCATCAAGGAAACCCAGAAACACTGAAACACTC 2040  
 Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
 QY 2041 TTCTGTACTTTTGAAGAGTCAAGAGAAAGGAGTGTGAAAGGAGTGTGAGATTTGACATTA 2100  
 Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
 QY 2101 CAGAGTCTTGAAGATTTCAGAAATGGGAAAGGTTGTGAAAGGCTCTGACACGCTTACAC 2160  
 Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
 QY 2161 GTCACCTTACGAGGTACCATAGAGGCAACGCGCGAGGCGATGCTACAGTGGATTTTGCA 2220  
 Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
 QY 2221 AATCGTTTTTGTGGAGGTGGTGTGACTGCTGGTGGGAGCTTGTACAGAAAGAAATCAGATT 2280



Db 740 AsnArgpheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluLeuArgPhe 759

QY 2281 TTAATCAATCTGAATGATTGTTTCACGGCTGTTCACTGAGGTGCTGATCAACAATGAG 2340

Db 760 LeuileAsnProGluLeuileileSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTACAGTACAGGCTATGCTGAAACTTAT 2400

Db 780 CysLeuileileThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799

QY 2401 CGTTGGCCCCAAGCATGAAGATGGAGTGAAGAGGAGGATGGCAGCGCGCTGCACG 2460

Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThr 819

QY 2461 GAGATCGTTGCCATTGACGCACTTCACTTCAGCGCTACCTCGATCAGTTTGTGCTGAG 2520

Db 820 GluileValAlaileAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

QY 2521 AAAGTGAGCGTGAGCTTAACAAGGCTTACTGCGGATTCCTCGCTCCTGAGTTCTTCT 2580

Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

QY 2581 GAAATCTTTCTGAGTGGCCACGGGAACTGGGCTGTGGTGCCTTTGGGGTGACGCT 2640

Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879

QY 2641 AGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGGGCTGAACGTGACGTGTTTAT 2700

Db 880 ArgLeuLysAlaLeuileGlnileLeuAlaAlaAlaAlaGluArgAspValValTyr 899

QY 2701 TTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGACACTTTCCTTACC 2760

Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

QY 2761 GAGAGGAAGCTGGATGTTGGAAGAGTGTACAAGTATTGCTTAGATACACTACATGAAGA 2820

Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuArgTyrTyrAsnGluGlu 939

QY 2821 TGCAGAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 2880

Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

QY 2881 GTTGAGTCAAGTGCAAGACCACTGACATGCCAGACAGAAGGCAGGCACC 2931

Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 9

AAE25629

ID AAE25629 standard; protein; 977 AA.

XX AAE25629;

AC AAE25629;

XX

DT 04-NOV-2002 (first entry)

XX

DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).

XX

KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;

KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;

KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;

KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;

KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;

antisense therapy.

XX

OS Bos taurus.

XX

PN US6395543-B1.

XX

PD 28-MAY-2002.

XX

XX 23-FEB-2000; 2000US-00511507.

PF

XX 01-MAY-1998; 98US-0083768P.

PR

PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

PA

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

PI

XX WPI; 2002-535641/57.

DR N-PSDB; AAD42081.

DR

XX

PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

PT

XX Claim 3; Col 47-45; 77pp; English.

PS

XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyzes release of ADP-ribose from an ADP ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The antisense therapeutic are useful to treat neoplastic disorders and conditions caused by genotoxic oxidative stress e.g., cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. The invention is useful in gene therapy and antisense therapy. The present sequence is bPARG

XX

SQ Sequence 977 AA;

Alignment Scores:

Pred. No.: 0 Length: 977

Score: 4332.50 Matches: 820

Percent Similarity: 89.86% Conservative: 57

Best Local Similarity: 84.02% Mismatches: 90

Query Match: 63.97% Indels: 9

DB: 5 Gaps: 5

US-09-302-812-5 (1-3814) x AAE25629 (1-977)

QY 28 ATGAGTGGGGCCCGGCTGGGAGCCCTGCACGAAA--GCGCGTGGGGCGCGCTGGA 84

Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20

QY 85 ACTTCTGCGCGGACTGCTCGGACTCCGGAGCTTCCCTGGCAGGCGAGCGTGTCTC 144

Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40

QY 145 GACCCCAAGGACGCTCCGCTCCAGTTCAGGTCCCTCCGCTCCGCGAGCTGCTCTCG 204

Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

QY 205 GGGCGGGCGGACCGCACAGAGGCAACGCCACCTCGTTGTTTTCACAAAAGACTATT 264

Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

QY 265 ACTACTTGGATGATATACTAAAGGACCCAGACAGCTGAATCAGAA-----AGTAAA 315

Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

QY 316 GAAACAACAATACAAAGATTGACTCCATGATGAGTTCTGTGAGAAAGATAACTTTTAC 375

Db 101 GluAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

QY 376 CCACATAAGCTGAAAAAATTGGAAAAATGTTCTCTCAGCTAAATCTTGATAAATCACCACA 435

Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

QY 436 GAAAAGAGTTCACAGTATTGTAACCAACACAGCAGCTGCGAGTGTGTGCAAGTGCAGAA 495

Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

QY 496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCGCTCCCGGGGAGCTCCGCTACCA 555  
 Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
 QY 556 AAGCAGCTTAGTAATGCTAAACATTGGTCAGTCACTCCACACACTGATGACCACTGACACA 615  
 Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspHisSerAspThr 200  
 QY 616 GATCATGAAGAAGACAGACAAATCAGCAGTTTCTTACACCTATAAACTGCAAAATACA 675  
 Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
 QY 676 AAGCCAAACAGTAGGAGATGGGACAG-----GCCAGAAGCAACTGTAAGTGCAGTGA 726  
 Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
 QY 727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGTGTGCTACCA 786  
 Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
 QY 787 GAGAGTCTTTGTTCAGATGTTGGTCCGAGGACATTGGAACCTGGACCAAAAATGACAAC 846  
 Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 QY 847 AAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCCTCCATTGAGAAAGAAAGTGAG 906  
 Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProPheGluLysGluSerGlu 300  
 QY 907 CCTGAGTCAACCAATGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAAGAGCAGAT 966  
 Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
 QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCTCTCC---CAAACAGCAAAT 1023  
 Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerAlaGlnThrAlaAsn 340  
 QY 1024 AAACCTTCAAGTTGCCAAGCAAGAGAGCTGATGCGCATCTTAGGAAACGGTATTGACT 1083  
 Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
 QY 1084 AAGGGAAGTGAAGTATGATTTGCATTTTCCAAATTTGAA--GGAGAAATAATGCTGGGACC 1140  
 Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
 QY 1141 AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTAGCCTTAATGTAGAGTGTAGAGT 1200  
 Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
 QY 1201 TCCAAGCAGCATGGAAGAAAGGATCTCTAAATTACAGATCATTTCTATGAGAAATTTCCAAG 1260  
 Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
 QY 1261 TCAGAGGACAGAAAGAAACAATGTGAAGTCAAGATCAAGATCAATGAGAAATTTCCAAG 1320  
 Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
 QY 1321 CCAAAATACATCCCACTTAACCTCCCTCCAGAGAAGAAGTGGTGGAACTCCTATTGAG 1380  
 Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
 QY 1381 GAAATGAGAAAAATGCCCTCGGTGTGGATCCATTTGCTTCCCTTAAGACCATCTGCAAGT 1440  
 Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
 QY 1441 CACACAGTGACTGTTCCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500  
 Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
 QY 1501 ACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAAATGCTTGTTCGGAACAAAAC 1560  
 Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
 QY 1561 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGCAACTGCAGGGAGTAGGTGGGAGCTCATT 1620

Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
 QY 1621 CAGACTGCATTCTCAACAAATTCACACGACCCAGAACTTGAAGGATGCGATTCTGAAA 1680  
 Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
 QY 1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTGTTGATTCTGGGATAAG 1740  
 Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
 QY 1741 GTACTTGAAGAAGCAGAGGCCCCAACATTTATATATATATATATATATATATATATAT 1800  
 Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
 QY 1801 ATTGCACCTCTGTCTGCCAAATATTTGCACCCAGCCCAATACCACTCTCCTGAACAGATG 1860  
 Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
 QY 1861 AATCATTTCTGTCACGATGTACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTCTTC 1920  
 Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
 QY 1921 TGCACATTTCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 1980  
 Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
 QY 1981 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACCCAGAAAACTGAAAAACACTC 2040  
 Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
 QY 2041 TTCTGCTACTTTTCGAAGAGTCCAGAGAACAGAAAAAACCTTACAGGATTGGTGACATTTACAAGA 2100  
 Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
 QY 2101 CAGAGTCTTGAAGATTTTCCAGAAATGGGAAGCTGTGTGAAAGCCTCTGACACGCTTACAC 2160  
 Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
 QY 2161 GTCACCTTACGAGGTACCATAGAAGGCAACCGCCGAGGCATGCTACAGTGGATTGCA 2220  
 Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
 QY 2221 AATCGTTTGTGGAGGTGGTGTGACTGTGTCGGGACTTGTACAAGAAATAATCAGATT 2280  
 Db 741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
 QY 2281 TTAATCAATCCTGAATTGATTGTTTTCACGGTGTTCACAGGTGTTCAGTGGTGTGATCACAATGAG 2340  
 Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
 QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAAACTTAT 2400  
 Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 QY 2401 CGTTGGGCCCCGAAGCCATGAAGATGGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACG 2460  
 Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820  
 QY 2461 GAGATCGTTGCTTACGCACTTCACTTACAGCAGCTACCTCGATCAGTTTGTGCTGAG 2520  
 Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
 QY 2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGCTCCTGGAGTTCTTCT 2580  
 Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 QY 2581 GAAATCTTTCTGCAGTGGCCACGGGAAACTGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640  
 Db 861 GluAsnLeuSerAlaValaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
 QY 2641 AGATTAAAGCCCTTAATACAGATCCTGCGCAGCTGCTGCGGCTGAACGTCGCTGTTAT 2700

Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
Qy 2701 TTCACCTTTGGGACTCAGAGTTGTATGAGAGACATTTACAGCATGACACTTTCTCTTACC 2760  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
Qy 2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAAGTTATTGCTTAGTACTACATGAAGAA 2820  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
Qy 2821 TGCAGAAACTGTTCCACCCCTGGACCATCAAGCTTTTATCCATTCATATACCATGCT 2880  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
Qy 2881 GTTGAGTCAAGTCAGAGACCATGACATGCCAGGACAGAGGAGGC 2928  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

RESULT 10

AAU76020  
ID AAU76020 standard; protein; 977 AA.

XX AC AAU76020;

DT 08-MAY-2002 (first entry)

DE Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX KW Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.

XX OS Bos taurus.

XX FH Key Location/Qualifiers

FT Region 501..617  
FT /note= "Represents PARG oligopeptide #1"  
FT Region 761..770  
FT /note= "Represents PARG oligopeptide #2"  
FT Region 771..801  
FT /note= "Represents PARG oligopeptide #3"  
FT Region 849..880  
FT /note= "Represents PARG oligopeptide #4"

XX PN US6337202-B1.

XX PD 08-JAN-2002.

XX PF 23-FEB-2000; 2000US-00511477.

XX PR 01-MAY-1998; 98US-0083768P.

XX PR 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;

XX DR WPI; 2002-163240/21.

XX DR N-PSDB; ABK14931.

XX PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.

XX PS Claim 2; Col 47-52; 81pp; English.

XX CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or

CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the bovine PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention

XX SQ Sequence 977 AA;

Alignment Scores:

Pred. No.: 0 Length: 977  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 63.97% Indels: 9  
DB: 5 Gaps: 5

US-09-302-812-5 (1-3814) x AAU76020 (1-977)

QY 28 ATGAGTCCGGGCCCCCGCTGGAGCCCTGCACGAAA--GCGCGCTGGGCGCGCTGGA 84  
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
QY 85 ACTTCTGCGCGGACTGCTCCGACTCCGGAGCTTCCCTGGCAGGCGAGCGTCTCTC 144  
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
QY 145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGTCCTCCCTCCGTCCTGCCAGCCTGCTCTCG 204  
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerGlyCysAlaLeu 60  
QY 205 GGGCGGGCGGACCGCACAGAGGCAACGCCACCTCGTTTGTTCATAACAAAGACTATT 264  
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrile 80  
QY 265 ACTACTTGATGGATACTAAAGGACCCCAAGACAGCTGAATCAGAA-----AGTAAA 315  
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
QY 316 GAAACACAAATACAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTAC 375  
Db 101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120  
QY 376 CCACATAAGGTGGAAAAATTGGAAAATGTTCTCTCAGCTAAATCTTGATAAATCACCACA 435  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 436 GAAAGAGATTACAGTATTGTAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495  
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCTCCCGCGGGACTCCGCTACCA 555  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 556 AAGCAGCTTAGTAATGCTAACATTGTTGTCAGTCAACCCACACTGATGACACAGTGACACA 615  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
QY 616 GATCATGAAGAAGACAGACACAATCAGAGTTTCTTACACCTATATAAACTTGCAATACA 675  
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 676 AAGCCCAACAGTAGGAGATGGGCAG-----GCCAGAAGCAACTGTAAAGTCAGTGA 726  
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGATGTCTACCA 786



Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
 QY 787 GAGAGTCCCTTTGTCAGATGTTGGTCCGAGGACATTGGAACCTGGACCAAAAATGACAAC 846  
 Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 QY 847 AAATTGACTGGACAAGAACGACCTAGGTGATTGCGCTCCATTGAGAGAAAGAGTGAG 906  
 Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProPheGluLysGluSerGlu 300  
 QY 907 CCTGAGTCACCAATGATGTAGACAACCTCGAGAAACAGTTGTCAAGATTGAGAACGAGAT 966  
 Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
 QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCTCTCC---CAAACAGCAAAT 1023  
 Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerAlaGlnThrAlaAsn 340  
 QY 1024 AAACTTTCAAGTTGCCAAGCAAGAGAGCTGATGCGGATCTTAGGAAACGTFATTGACT 1083  
 Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
 QY 1084 AAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAA---GGAGAAATAATGCTGGGACC 1140  
 Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
 QY 1141 AGTGACTTAAATGCCAAGCCATCTCTGGAACCTCTTAGCCCTTAATGTAGAGTGTAGAAGT 1200  
 Db 381 AsnAspValAsnAlaLysArgProGlySerSerSerLeuAsnValGluCysArgAsn 400  
 QY 1201 TCCAAGCAGCATGGAAGGATCTCTAAATATACAGATCATTTTCATGAGAAATTTCCAAG 1260  
 Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
 QY 1261 TCAGAGGACAGAAAGAAAGCAATGTGAAGTCAGACATCAAGAACAGAAAGCAAGATT 1320  
 Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
 QY 1321 CCAAAATACATCCCACTTAACCTCCCTCCAGAGAGAGAGTGGCTGGGAACCTCTATTGAG 1380  
 Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
 QY 1381 GAAATGAGAAAAATGCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGT 1440  
 Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
 QY 1441 CACACAGTGACTGTTGCGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500  
 Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
 QY 1501 ACACATTACAAAGATTGTTGGGATAACAAACATGTGAAAAATGCCTTGTTCGGAACAAAAC 1560  
 Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
 QY 1561 TTGTACCTGTGGAAGATGAGAATGGTGAGCGGAACCTGCAGGAGTAGGTGGGAGCTCATT 1620  
 Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
 QY 1621 CAGACTGCATTTCTCAACAAATTCACACGACCCAGAACCTTGAAGGATGCGATTCTGAAA 1680  
 Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
 QY 1681 TACAAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGTTGATTCTCGGATAAG 1740  
 Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
 QY 1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAG 1800  
 Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
 QY 1801 ATTGCACCTCTGTCTGCCAAATATTGACCCAGCCCAATACCACCTCTCTGAAACACAAGATG 1860

Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
 QY 1861 AATCATTCTGTACGATGTACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920  
 Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
 QY 1921 TGCACATTTCCTCCGACGGAATGCCAAGATGCCAATCGGAGTATTCTAGTTACCCAGACATT 1980  
 Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
 QY 1981 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACCCAGAAAACTGAAAAACACTC 2040  
 Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
 QY 2041 TTCTGTCTACTTTCGAAGAGTCCAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGA 2100  
 Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
 QY 2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACAC 2160  
 Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
 QY 2161 GTCACATTACGAGGTACCATAGAAGGCAACCGCCGAGGCATGCTACAGGTGATTTTGCA 2220  
 Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
 QY 2221 AATCGTTTTGTGGAGGTGTTGACTGCTGGGGACTTGTACAAGAAAGAAATCAGATTT 2280  
 Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
 QY 2281 TTAATCAATCCTGAATGATTGTTTTCACGCGCTGTTCACTGAGGTGCTGGATCACAATGAG 2340  
 Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
 QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400  
 Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 QY 2401 CGTTGGGCCCGAAGCCCATGAAGATGGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACG 2460  
 Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820  
 QY 2461 GAGATCGTTGCCATTGACGCATCTCATTACAGACGCTACCTCGATCAGTTGTGCTTCT 2520  
 Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
 QY 2521 AAAGTGAGACGCTGAGCTTAACAAGGCTTACTCGCGGATTCCTCCGCTCCTGGAGTCTTCT 2580  
 Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 QY 2581 GAAAATCTTTCTCAGTGGCCACCGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640  
 Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
 QY 2641 AGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGGGCTGAACGCTGAGCTGGTTTAT 2700  
 Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
 QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTCTTACC 2760  
 Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
 QY 2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAAGTTATTGCTTAGATACATAATGAAGAA 2820  
 Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
 QY 2821 TGCAGAAAACCTGTTCCACCCCTGGACCCAGACATCAAGCTTTTATCCATTATATACCATGCT 2880  
 Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
 QY 2881 GTTGAGTCAAGTGCAGAGACCACCTGACATGCCAGGACAGAGGAGGAGGAGGAGGAGG 2928  
 Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976





Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
QY 1084 AAGGGAAGTGAAGTTAGATTGCAATTTCCAATTTGAA---GGAGAAAATAATGCTGGGACC 1140  
Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
QY 1141 AGTGACTTAAATGCCAAGCCATCTCTGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200  
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
QY 1201 TCCAAGCAGCATGGAAAAGGGATTCTAAATTTACAGATCATTTTCATGAGAATTTCCAAG 1260  
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
QY 1261 TCAGAGGACAGAAAGAAAACAATGTGAAGTCAGACATCAAAAGAACAGAAAGGAAGATT 1320  
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
QY 1321 CCAAAATACATCCCACTAACCTCCCTCCAGAGAGAAGTGGCTGGAACTCCTATTGAG 1380  
Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
QY 1381 GAAATGAGAAAATGCCCTCGGTGTGGGATCCATTTGCCCTTCAAGACCATCTGCAAGT 1440  
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
QY 1441 CACACAGTGACTGTTGCGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500  
Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
QY 1501 ACACATTACAAAGATTGTTGGGATAACAACATGTGAATGCTTGTTCGGAAACAAAC 1560  
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
QY 1561 TTGTACCCTGTGGAAGATGAAATGGTGAGCGAACTGACGGAGTAGGTGGGAGCTCAT 1620  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
QY 1621 CAGACTGCACCTTCTCAACAATTCACACGACCCCAAGACTTGAAGGATGCGATTCTGAAA 1680  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
QY 1681 TACAATGTGGCATATTTCAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740  
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
QY 1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAG 1800  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
QY 1801 ATTGCACCTCTGTCTGCCAATATTTGCACCCAGCCCAATACCCTCTGAAACAGAGATG 1860  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
QY 1861 AATCATTTCTGCATGTACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
QY 1921 TGCACATTTCCCGACGGAATGCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 1980  
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
QY 1981 AACTTCAATCGGTTGTTGAAGACGTTTCATCAAGGAACCAAGAAAACCTGAAACACTC 2040  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
QY 2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGA 2100  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2101 CAGAGTCTTGAAGATTTTCAGAAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACAC 2160  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

QY 2161 GTCACCTTACGAGGGTACCATAGAAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTTCGA 2220  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2221 AATCGTTTGTGGAGGTGTGTGACTGGTGGGAGCTTGTACAAGAAATCAGATT 2280  
Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
QY 2281 TTAATCAATCCTGAATGATTTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAG 2340  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2341 TGTCTTATTATCACAGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAACTTAT 2400  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2401 CGTTGGCCCGAAGCCATGAAGATGGGAGTGAAGAGACGATTGGCAGCGCGCTGCACG 2460  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgArgThrThr 820  
QY 2461 GAGATCGTTGCCATTGACGCACCTTCACCTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2520  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2521 AAAGTGAGACGTGAGCTTAAAGGCTTACTGCGGATTCCTCCGCTCGGAGTTCTTCT 2580  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2581 GAAATCTTTCTGAGTGGCCACCGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
QY 2641 AGATTAAGCCCTTATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGTTTAT 2700  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 2761 GAGAGGAAGCTGGATGTTGAAAAGTGTACAAGTTATTGCTTAGATACATAATGAAGAA 2820  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
QY 2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 2880  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAGGCAGGC 2928  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

RESULT 12

ABG72278

ID ABG72278 standard; protein; 977 AA.

XX AC ABG72278;

XX DT 13-MAR-2003 (first entry)

XX DE Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.

XX KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
cellular response; DNA damage; neoplastic disorder inducing agent;  
oxidative stress; neoplastic disorder; myocardial infarction;  
vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
Parkinson's disease; Huntington's disease; inborn genetic error;  
reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;  
anticonvulsant; cerebroprotective; enzyme.

OS Bos taurus.

XX US2002132328-A1.  
PN 19-SEP-2002.  
XX 09-OCT-2001; 2001US-00973451.  
XX 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2003-155895/15.  
DR N-PSDB; ABX14477.  
XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX Claim 28; Fig 16; 86pp; English.  
XX The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents bovine PARG enzyme  
XX Sequence 977 AA;  
SQ  
Alignment Scores:  
Pred. No.: 0 Length: 977  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 63.97% Indels: 9  
DB: 6 Gaps: 5  
US-09-302-812-5 (1-3814) x ABG72278 (1-977)  
QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCACGAAA---GCGCGCTGGGGCGCGCTGGA 84  
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
QY 85 ACTTCTGCGCGACTGCTCGGACTCCCGAGCTTCCCTGGCAGGAGCGGTGTTCTC 144  
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
QY 145 GACCCCAAGGACCTCCCGTCCAGTTCAGGTCCCTCCGTCCCTGCCAGCCTGCTCG 204  
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
QY 205 GGGCGGGGACCGCACAGAGGCAACGCCACCTCGTTGTTTCAACAAAGACTATT 264  
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
QY 265 ACTACTTGGATGGATATACTAAAGGACCAAGACAGCTGAATCAGAA-----AGTAA 315

Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
QY 316 GAAAAACAATACAAAGAAATTGACTCCATGATGAGTCTTGTGCAGAAAGATACTTTTAC 375  
Db 101 GluAsnAsnThrArgGluSerMetMetSerValGlnLysAspAsnPheTyr 120  
QY 376 CCACATAAGGTGGAAAAATTGGAAAAATGTTCTCTCAGCTAAATCTTGATAAATCACCACA 435  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 436 GAAAAGAGTTCACAGTATTGAAACCAACACGACGACTGCGAGTGTGTGCAAGTGGCAGAAT 495  
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 496 GAAGGGAAGCATGCAGAACACAGCTTTTGGCAAGTGAAGTCCCGGGGAGCTCCGCTACCA 555  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 556 AAGCAGCTTAGTAATGCTAACATTGGTCACTCACCACACACTGATGACACAGTGACACA 615  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
QY 616 GATCATGAAGACAGACAGACAATCAGCAGCTTCTTACACCTATATAAACTTGCAATACA 675  
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 676 AAGCCCAACAGTAGAGATGGGCAG-----GCCAGAAGCAACTGTAAGTGCAGTGA 726  
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTCTACCA 786  
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValSer 260  
QY 787 GAGAGTCTTTGTGATGTTGGTGGCGAGGACATGGAACCTGGACCAAAAATGACAAC 846  
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
QY 847 AAATTGACTGGACAAGAACAGCAGCTTAGTGTATTCGCTCCATTTGAGAAAGAAAGTGAG 906  
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
QY 907 CCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTGTCTCAAGATTGAGAAGCAGAT 966  
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCTCTCC---CAACACAGCAAAT 1023  
Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
QY 1024 AAACCTTCAAGTTGCCAAGCAAGAGAGCTGATGGCGATCTTAGGAAACGGTATTGACT 1083  
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
QY 1084 AAGGGAAGTGAAGTAGATTGATTTCCATTTCCAAATTTGAA---GGAGAAAATAATGCTGGGACC 1140  
Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
QY 1141 AGTGACTTAAATGCCAAGCCATCTGGAACCTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200  
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
QY 1201 TCCAGCAGCATGGAAAAAGGATTTCTAAATTTACAGATCATTTTCATGAGAAATTTCCAGG 1260  
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
QY 1261 TCAGAGCAGACAGAAAGAACAAATGTGAAGTCCAGACATCAAGAAACAGAAAGGAGATT 1320  
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
QY 1321 CCAAAATACATCCCACCTAACCTCCCTCCCTCCAGAGAAAGAGTGGTGGGAACCTCTATTGAG 1380



Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
QY 1381 GAAATGAGAAAAATGCCTCGGTGTGGGATCCATTGTCCTTCCCTTAAGACCATTCTGCAAGT 1440  
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
QY 1441 CACACAGTGAAGTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCGGAACCTTTTCCA 1500  
Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
QY 1501 ACACATTACAAAGATTGTGGGATAACAAACATGTGAAATGCCTTGTTCGGACAAAC 1560  
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
QY 1561 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGAGCTCATT 1620  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
QY 1621 CAGACTGCACTTCTCAACAATTCACACGACCCCGAAGTGAAGATCGGATTCGAA 1680  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
QY 1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTCTGGGATAAG 1740  
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
QY 1741 GTACTTGAAGAGCAGAGGCCCAACATTATATACATGTCATTTTACCTGACATGGTGA 1800  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
QY 1801 ATTGCACCTGTCTGCCAAATATTGTCACCCAGCCCAATACCACCTCTCTGAAACAGAGATG 1860  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
QY 1861 AATCATCTGTCTCAGCATGTACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
QY 1921 TGCACATTTCCCGGACCGAATGCCAAGATGAATCGGAGTATTCTAGTTACCCAGACATT 1980  
Db 641 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerLysProAspIle 660  
QY 1981 AACTTCAATCGTTTGAAGGACGTTTCATCAAGAAACCCAGAAACCTGAAACACACTC 2040  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
QY 2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCCTACAGGATTGGTGACATTTACAAGA 2100  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2101 CAGAGTCTTGAAATTTCCAGAATGGGAAAGGTGTGAAAGCCCTCTGACACGCTTACAC 2160  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
QY 2161 GTCACCTTACGAGGTACCATAGAAAGCAACGGCCGAGGCATGCTACAGGTGATTTTGA 2220  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2221 AATCGTTTTTGTGGAGGTGGTGTGACTGGTGGGACTGTGTACAAAGAAATCAGATTT 2280  
Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
QY 2281 TTAATCAATCCTGAATTGATTGTTTACGGCTGTTCACTGAGGTGCTGATCACAATGAG 2340  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2341 TGTCTTATTATCAGAGTACTGAACAGTACAGTACAGTAAATACACAGGCTATGCTGAAACTTAT 2400  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2401 CGTTGGCCCGAAGCCATGAAGATGGGAGTGAAGACGATTGGCAGCGCGCTGCACG 2460  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgThrThr 820

QY 2461 GAGATCGTTGCCATTGACGCACTTCACTTACAGACGCTACCTCGATCAGTTTGTGCCTGAG 2520  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2521 AAAGTGAGACGCTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGTCCTCGGAGTTCTTCT 2580  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2581 GAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
QY 2641 AGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGGCTGAACGTGACGTGCTTAT 2700  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 2761 GAGAGGAAGCTGGATGTTGAAAAGTGTAACAGTTATTGCTTAGATACATCAATGAAGAA 2820  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
QY 2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTATATACCATGCT 2880  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 2881 GTTGAGTCAAGTGACAGACCACTGACATGCCAGGACAGAAAGGCAGGC 2928  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976  
RESULT 13  
ABB59491  
ID ABB59491 standard; protein; 768 AA.  
XX  
AC ABB59491;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 5265.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WC200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL03594.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
PS Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and







QY 2467 GTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTGTGCTGAGAAAGTG 2526  
Db ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet 471  
QY 2527 AGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGT-----CCT 2568  
Db GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProProPro 491  
QY 2569 GGAGTTCCTCTGAAAAATCTTCTGCAGTGGCCACGGAAACTGGGCTGTGCTGCTTT 2628  
Db Gly-----ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 503  
QY 2629 GGGGTGACGCTAGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGGCTGAACGT 2688  
Db GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523  
QY 2689 GACGTGGTTTATTTCACTTTGGGACTCAGAGTTGATGAGACATTTACAGCATGCAC 2748  
Db ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMetTrp 543  
QY 2749 ACTTTCCTTACCAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATGCTTAGATAC 2808  
Db LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSer 562  
QY 2809 TACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA-----GAC 2850  
Db TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582  
QY 2851 ATCAAGCTTTATCCATTTCATATACCATGCTGTTGAGTCAAGTGCAGAGACCACCTGACATG 2910  
Db LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599  
QY 2911 CCAGGACAGAGGCA 2925  
Db ProGlyGluGlyAla 604  
RESULT 15  
AAU76023  
ID AAU76023 standard; protein; 768 AA.  
XX AAU76023;  
AC AAU76023;  
XX 08-MAY-2002 (first entry)  
XX Fruit fly poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
DE Fruit fly; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
XX adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX Drosophila melanogaster.  
OS US6337202-B1.  
XX PN 08-JAN-2002.  
XX PF 23-FEB-2000; 2000US-005111477.  
XX PR 01-MAY-1998; 98US-0083768P.  
XX PR 30-APR-1999; 99US-00302812.  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX PI WPI; 2002-163240/21.  
XX DR N-PSDB; ABK14934.  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for

treating neoplastic and neurological disorders, heart attack and stroke.  
Example 13; Col 71-76; 81pp; English.  
The present invention relates to a new poly(ADP-ribose) glycohydrolase (PARG) protein which catalyses release of ADP-ribose from an ADP (adenosine diphosphate)-ribose polymer. The PARG molecule of the invention is useful for generating antibodies and can be inhibited or activated for diagnosing and treating neoplastic disorders such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, hyperplasia and hypertrophy, reperfusion following ischaemia, heart attack, stroke, neurodegenerative diseases, neurological disorders including Alzheimer's, Huntington's and Parkinson's diseases, and related conditions. PARG levels may be enhanced to suppress DNA repair and increase the cell's susceptibility to chemotherapy drugs. Antagonists of PARG are administered to treat or prevent neoplastic disorders. The present amino acid sequence represents the fruit fly PARG protein of the invention. This protein is one of several PARG proteins (AAU76020-AAU76024) of the invention  
XX SQ Sequence 768 AA;  
Alignment Scores:  
Pred. No.: 1.72e-93 Length: 768  
Score: 1074.50 Matches: 244  
Percent Similarity: 58.12% Conservative: 96  
Best Local Similarity: 41.71% Mismatches: 194  
Query Match: 15.86% Indels: 51  
DB: 5 Gaps: 15  
US-09-302-812-5 (1-3814) x AAU76023 (1-768)  
QY 1249 AGAATTTCCAGTACAGGACAGAGAAAGAACAAATGTGAAGTCAGACATCAAGAACA 1308  
Db 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGlu----- 62  
QY 1309 GAAAGGAAGATTCCAAATACATCCACCTAACCTCCCTCCA-----GAGAGAAGTGG 1362  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTrp 74  
QY 1363 CTGGGAACCTCTATTGAGGAAATGAGAAAATGCTCGGTGGGATCCATTTGCTTCC 1422  
Db 75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93  
QY 1423 TTAAGACCA---TCTGCAAGTCACACAGTCACTGTTGCGGTAGACCTTCTGAGAGCAGGA 1479  
Db 94 LeuProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGlu 113  
QY 1480 GAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTTGTGGGATAACAACATGTGAAA 1539  
Db 114 ThrProProArgProTyrLysSerProGlyLys-----TrpAspSerGluHisValArg 131  
QY 1540 ATGCTTGTTCGGAACAAACTGTACCTGTGGAGATGAGAATGTTGAGCGAACTGCA 1599  
Db 132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrile 151  
QY 1600 GGGAGTAGGTGGGAGCTCATTCAGACTGCATCTCTCAACAAATTCACACGCCCCAGAAC 1659  
Db 152 AspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171  
QY 1660 TTGAAGGATCGGATTTCTGAAATACAAATGTGGCATATTTAAGAAATGGACCTTTACAGCT 1719  
Db 172 LeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAla 191  
QY 1720 TTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACATTATATCAGTCC 1779  
Db 192 LeuHisGlnLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211  
QY 1780 ATTTTACCTGACATGGTGAATTCACACTCTGTCTGCCAAATATTTGCCACCCAGCCAATA 1839  
Db 212 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerProVal 231  
QY 1840 CCACCTCTGAAACAGAGATGAATCATCTGTCTCAGATGTCTACAGGAACAGATCGCCAGT 1899

|    |      |  |      |
|----|------|--|------|
| Db | 232  | proLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnSerCys      | 251  |
| QY | 1900 | CTTTTAGCTAATGCTTTCTTCTGCACATTTCCCGACGGAATGCC---                | 1956 |
| Db | 252  | LeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSer   | 271  |
| QY | 1957 | GAGTATTCTAGTTACCCAGACATTAACCTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGG | 2016 |
| Db | 272  | GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla   | 291  |
| QY | 2017 | AAACCAGAAAAAAGCTGAAAAACACTCTTCTGCTACTTTTGAAGAGTC-----ACAGAGAAA | 2070 |
| Db | 292  | ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg   | 311  |
| QY | 2071 | AAA-----CCTACAGGATTGGTGACATTTACAAGACAGAGT-----CTTGAA           | 2112 |
| Db | 312  | AspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGlu   | 331  |
| QY | 2113 | GATTTTCCAGAATGGAAAGGTGTGAAAAGCCCTCG-----ACAGCTTACACGTCAC       | 2166 |
| Db | 332  | HisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp   | 351  |
| QY | 2167 | TACGAGGGTACCATAGAAGCCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAAATCGT  | 2226 |
| Db | 352  | AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys   | 371  |
| QY | 2227 | TTTGTGGAGGTGGTGACTGGTGGCGGACTTGACAAGAAGAAATCAGATTTTAAATC       | 2286 |
| Db | 372  | TyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheValIle   | 391  |
| QY | 2287 | AATCCTGAATTGATTGTTTACGGCTGTTCACAGGTGCTGGATCAACAATGAGTGTCTT     | 2346 |
| Db | 392  | CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu   | 411  |
| QY | 2347 | ATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGG   | 2406 |
| Db | 412  | ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTrp   | 431  |
| QY | 2407 | GCCCCAAGCCATGAAGATGGAGTGAAGAAGACGATGGCAGCGCGCTGCACGAGATC       | 2466 |
| Db | 432  | SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgGlnThrAlaIle      | 451  |
| QY | 2467 | GTTGCCATTGACGCATTCACTTCAGACGCTACCTCGATCAGTTTGTGCTGAGAAAGTG     | 2526 |
| Db | 452  | ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet   | 471  |
| QY | 2527 | AGACGTGAGCTTAACAAGGCTTACTGCGGATTCTCCGT-----CCT                 | 2568 |
| Db | 472  | GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro      | 491  |
| QY | 2569 | GGAGTTCCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTCTTT    | 2628 |
| Db | 492  | Gly-----ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe                      | 503  |
| QY | 2629 | GGGGGTGACGCTAGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTCGCGGTGAACGT    | 2688 |
| Db | 504  | GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg   | 523  |
| QY | 2689 | GACGTGGTTTATTTACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCAC     | 2748 |
| Db | 524  | ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMetTrp   | 543  |
| QY | 2749 | ACTTTCCTTACCGAGAGGAAGCTGGATGTTTGGAAAAGTGTACAAAGTTATTGCTTAGATAC | 2808 |
| Db | 544  | LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSer   | 562  |
| QY | 2809 | TACAATGAAGAATGCAGAAACTGTTCACCCCTGGACCA-----GAC                 | 2850 |
| Db | 563  | TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys   | 582  |
| QY | 2851 | ATCAGCTTTTATCCATTTCATATACCATGCTGTTGAGTCAAGTCAGAGACCACCTGACATG  | 2910 |

Db 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysValArgAspVal 599

QY 2911 CCAGGACAGAAGGCA 2925  
|||||:::||||

db 600 proGlyGluGlyAla 604

Search completed: May 26, 2004, 16:49:22  
Job time : 274.747 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:27:48 ; Search time 31.9083 Seconds  
(without alignments)  
12341.705 Million cell updates/sec

Title: US-09-302-812-5  
Perfect score: 6773  
Sequence: 1 gggggactgtgtgctgcggg.....aatcatttgcagaaaaaa 3814

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xl  
-Q=/cgn2\_1/USPTO\_spool\_p/6333148/runat\_26052004\_150054\_5717/app\_query.fasta\_1.12437  
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148 @CGN 1 1 131 @runat\_26052004\_150054\_5717 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description     |
|------------|--------|-------------|--------|----|-----------------|
| 1          | 5164   | 76.2        | 968    | 4  | US-09-302-812-6 |
| 2          | 5164   | 76.2        | 968    | 4  | US-09-511-477-6 |
| 3          | 5164   | 76.2        | 968    | 4  | US-09-511-507-6 |
| 4          | 4395   | 64.9        | 976    | 4  | US-09-302-812-4 |
| 5          | 4395   | 64.9        | 976    | 4  | US-09-511-477-4 |
| 6          | 4395   | 64.9        | 976    | 4  | US-09-511-507-4 |
| 7          | 4332.5 | 64.0        | 977    | 4  | US-09-302-812-2 |
| 8          | 4332.5 | 64.0        | 977    | 4  | US-09-511-477-2 |
| 9          | 4332.5 | 64.0        | 977    | 4  | US-09-511-507-2 |
| 10         | 1074.5 | 15.9        | 768    | 4  | US-09-302-812-8 |
| 11         | 1074.5 | 15.9        | 768    | 4  | US-09-511-477-8 |
| 12         | 1074.5 | 15.9        | 768    | 4  | US-09-511-507-8 |

|    |       |     |      |   |                     |                    |
|----|-------|-----|------|---|---------------------|--------------------|
| 13 | 443.5 | 6.5 | 726  | 4 | US-09-302-812-10    | Sequence 10, Appl  |
| 14 | 443.5 | 6.5 | 726  | 4 | US-09-511-477-10    | Sequence 10, Appl  |
| 15 | 443.5 | 6.5 | 726  | 4 | US-09-511-507-10    | Sequence 10, Appl  |
| 16 | 167   | 2.5 | 31   | 4 | US-09-302-812-11    | Sequence 11, Appl  |
| 17 | 167   | 2.5 | 31   | 4 | US-09-511-477-11    | Sequence 11, Appl  |
| 18 | 167   | 2.5 | 31   | 4 | US-09-511-507-11    | Sequence 11, Appl  |
| 19 | 167   | 2.5 | 1177 | 4 | US-09-134-001C-5106 | Sequence 5106, Ap  |
| 20 | 166   | 2.5 | 1187 | 1 | US-08-320-559-28    | Sequence 28, Appl  |
| 21 | 166   | 2.5 | 1187 | 3 | US-08-545-860D-28   | Sequence 28, Appl  |
| 22 | 166   | 2.5 | 1187 | 5 | PCT-US94-04496-28   | Sequence 28, Appl  |
| 23 | 166   | 2.5 | 1210 | 1 | US-08-320-559-26    | Sequence 26, Appl  |
| 24 | 166   | 2.5 | 1210 | 3 | US-08-545-860D-26   | Sequence 26, Appl  |
| 25 | 166   | 2.5 | 1210 | 5 | PCT-US94-04496-26   | Sequence 26, Appl  |
| 26 | 157.5 | 2.3 | 559  | 1 | US-08-320-559-31    | Sequence 31, Appl  |
| 27 | 157.5 | 2.3 | 559  | 3 | US-08-545-860D-31   | Sequence 31, Appl  |
| 28 | 157.5 | 2.3 | 559  | 5 | PCT-US94-04496-31   | Sequence 31, Appl  |
| 29 | 155   | 2.3 | 1792 | 2 | US-08-962-284-4     | Sequence 4, Appli  |
| 30 | 152.5 | 2.3 | 1404 | 4 | US-08-801-308-1     | Sequence 1, Appli  |
| 31 | 151   | 2.2 | 913  | 4 | US-08-971-089-4     | Sequence 4, Appli  |
| 32 | 150   | 2.2 | 1115 | 2 | US-08-568-459A-2    | Sequence 2, Appli  |
| 33 | 150   | 2.2 | 1115 | 2 | US-08-487-826B-2    | Sequence 2, Appli  |
| 34 | 150   | 2.2 | 1115 | 4 | US-09-210-288-2     | Sequence 2, Appli  |
| 35 | 150   | 2.2 | 1115 | 6 | 5198347-6           | Patent No. 5198347 |
| 36 | 148.5 | 2.2 | 2329 | 3 | US-08-755-587-16    | Sequence 16, Appl  |
| 37 | 148   | 2.2 | 944  | 4 | US-09-134-001C-4352 | Sequence 4352, Ap  |
| 38 | 145.5 | 2.1 | 2442 | 4 | US-09-514-247A-10   | Sequence 10, Appl  |
| 39 | 145.5 | 2.1 | 3287 | 2 | US-08-477-451-7     | Sequence 7, Appli  |
| 40 | 144   | 2.1 | 26   | 4 | US-09-302-812-21    | Sequence 21, Appl  |
| 41 | 144   | 2.1 | 26   | 4 | US-09-511-477-21    | Sequence 21, Appl  |
| 42 | 144   | 2.1 | 26   | 4 | US-09-511-507-21    | Sequence 21, Appl  |
| 43 | 144   | 2.1 | 29   | 4 | US-09-302-812-12    | Sequence 12, Appl  |
| 44 | 144   | 2.1 | 29   | 4 | US-09-511-477-12    | Sequence 12, Appl  |
| 45 | 144   | 2.1 | 29   | 4 | US-09-511-507-12    | Sequence 12, Appl  |

ALIGNMENTS

RESULT 1

US-09-302-812-6  
; Sequence 6, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-302-812-6

Alignment Scores:  
Pred. No.: 0  
Score: 5164.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 76.24%  
DB: 4  
Length: 968  
Matches: 968  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-09-302-812-5 (1-3814) x US-09-302-812-6 (1-968)

QY 28 ATGAGTGGGGCCCCGGCTGGAGCCCTGCACGAAGCGCTGGGGCGCTGGGAACT 87

Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20  
QY 88 TCTGCGCGACTGCTCGAGACTCCGGAGCTTCCCTGGCAGGCAGAGCGTGTCTCGAC 147  
Db 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeuAsp 40  
QY 148 CCCAAGGACGCTCCCGTCAGTTCCAGGTCCTCCGTCCTCGCCAGCCTGCTCGGG 207  
Db 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
QY 208 CGGCGGACCGCACAGAGCAACGCCACCTCGTTTGTCTTCAACAAAGACTATTACT 267  
Db 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
QY 268 ACTTGATGGATACTAAAGGACCCAGACAGCTGAATCAGAAAGTAAGAAACAACAAT 327  
Db 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
QY 328 ACAAGAATTGACTCCATGATGAGTTCTGTGCGAGAAAGATAACTTTTACCCACATAAGGTG 387  
Db 101 ThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyrProHisLysVal 120  
QY 388 GAAAAATTGGAAATGTTCTCAGCTAAATCTTGATAAATCACCACAGAAAGAGTTCA 447  
Db 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer 140  
QY 448 CAGTATTGAACCAACACAGACACTGCGAGTGTGCAAGTGGCAGAATGAAGGAAGCAT 507  
Db 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160  
QY 508 GCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAGCAGCTTAGT 567  
Db 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
QY 568 AATGCTAACATTGGTCACTACCCACACACTGATGACCAACAGTGACACAGATCATGAAGAA 627  
Db 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200  
QY 628 GACAGAGACAATCAGCAGTTTCTTACACCTATAAACTTGCAATACAAAGCCAAACAGTA 687  
Db 201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220  
QY 688 GGAGATGGCGAGCCAGAACAACTGTAAAGTGCAGTGGATCTCGCCAGTCTGTGAAAAGAC 747  
Db 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240  
QY 748 TGTACAGGCTGTCAACACAGAGGAGGTGGATGTGCTACAGAGAGTCTTTGTCAGATGTT 807  
Db 241 CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal 260  
QY 808 GGTCCGAGGACATTGGAAGTGGACCAAAAAATGACAAATAATTGACTGGACAAAGAAC 867  
Db 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280  
QY 868 AGCCTAGGTGATTGCGCTCCATTGTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTA 927  
Db 281 SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal 300  
QY 928 GACAACTCGAGAAACAGTTGTCAAGATTGAGAGCAGATGAAGAAACAAGTCCAGTCTTT 987  
Db 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
QY 988 GATGAGCAAGATGATGTTCTCTCCAAACAGCAATAATAACTTTCAAGTTGCCAAGCAAGA 1047  
Db 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340  
QY 1048 GAAGCTGATGGCGATCTTAGAAACCGTATTGTAATAAGGAAGTGAAGTTAGATTGCAT 1107  
Db 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
QY 1108 TTCCAATTTGAAGGAGAAATAATGCTGGGACCACTGAATGATGCAAGCCATCTGGA 1167

Db 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGATTCT 1227  
Db 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
QY 1228 AAAATTACAGATCATTTTCATGAGAATTTCCAAGTCAGAGGACAGAGAAAGAAACAATGT 1287  
Db 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420  
QY 1288 GAAGTCAGACATCAAAAGAACAGAAAGAGAGATTCCAAAATACATCCCACCTAACCTCCCT 1347  
Db 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440  
QY 1348 CCAGAGAAGAGTGGCTGGGAACCTCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGG 1407  
Db 441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460  
QY 1408 ATCCATTGCTCTTCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCCGGGTAGACCTT 1467  
Db 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
QY 1468 CTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTGTGGGATAAC 1527  
Db 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500  
QY 1528 AAACATGTGAAAAATGCCTTGTTCGGAAACAAACTTGTACCCCTGTGGAAGATGAGAATGGT 1587  
Db 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
QY 1588 GAGCGAACTGACGGAGTAGGTGGAGTCATTCAGACTGCATCTTCAACAAATTCACA 1647  
Db 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuAsnLysPheThr 540  
QY 1648 CGACCCCGAAGCTTGAAGGATCGATTCTGAAATACAATGTGGCATATCTAAGAAATGG 1707  
Db 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560  
QY 1708 GACTTTACAGCTTTTGGTTGATTCTCGGATAAGGTACTTGAAGAAGCAGAGGCCCAACAT 1767  
Db 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluAlaGluAlaGlnHis 580  
QY 1768 TTATATCAGTCCATTTTACCTGACATGGTGAATAATGCACTCTGTCTGCCAAATATTTGC 1827  
Db 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCAGCCAAATACCACTCCTGAAACAGAAAGATGAATCATCTGTCTCAGATGTACAGGAA 1887  
Db 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATCGCCAGTCTTTTAGCTAATGCTTTTCTTCTGTCTGCACATTTCCCGCAGGAATGCCAAG 1947  
Db 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640  
QY 1948 ATGAAATCGGAGTATTCTAGTTACCCAGACATTAACCTTCAATCGGTGTTTGAAGGACGT 2007  
Db 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAAGGAAACAGAAAAAAGTGAACACACTTCTGTCTACTTTTCGAAGAGTCACAGAG 2067  
Db 661 SerSerArgLysProGluLysLeuLysThrLysThrLeuPheCysTyrPheArgArgValThrGlu 680  
QY 2068 AAAAACCTACAGGATTTGGTGACATTTTCAAGACAGAGTCTTGAAGATTTTCCAGAATGG 2127  
Db 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp 700  
QY 2128 GAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACCTTACGAGGGTACCATAGAAGGC 2187  
Db 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGCGCGAGGATGCTACAGGTGGATTTCGAAAATCGTTTTTGTGGAGGTGGTGTGACT 2247  
Db 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740

QY 2248 GGTGGGACTTGTACAAGAAATCAGATTTTAAATCAATCCTGAATGATGTTTCA 2307  
Db 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCAGAGTACTGAACAG 2367  
Db 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCCGAAGCCATGAAGATGGG 2427  
Db 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAGAAGGACGATTGGCAGCGGCTGCACGGAGATCGTTGCCATTGACGCACTTCAC 2487  
Db 801 SerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACCTCGATCAGTTTGTGCTGAGAAAGTGAGACGTGAGCTTAACAAGGCT 2547  
Db 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840  
QY 2548 TACTGCGGATTCTCTCCGCTGGAGTTCCTTCTGAAATCTTCTGAGTGGCCACGGGA 2607  
Db 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGCTGTGTGCTTTGGGGTGACGCTAGATTAAAGCCTTAATACAGATCCTG 2667  
Db 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
QY 2668 GCAGCTGCTGGGCTGAACGTGACGTGTTTATTCACCTTTGGGGACTCAGAGTTGATG 2727  
Db 881 AlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGAGACATTTACAGATGCACACTTTCTTACCGAGAGGAAGCTGGATGTTGGAAGAAGTG 2787  
Db 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
QY 2788 TACAAGTTATTGCTTAGATACTACAATGAAGATGAGAAACTGTTCCACCCCTGGACCA 2847  
Db 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTTATCCATTATATACCATGCTGTTGAGTCAAGTGCAGAGACCACTGAC 2907  
Db 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
QY 2908 ATGCCAGGACAGAGCGCAGCCACC 2931  
Db 961 MetProGlyGlnLysAlaGlyThr 968

RESULT 2

US-09-511-477-6  
; Sequence 6, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-511-477-6

Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 5164.00 Matches: 968  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 76.24% Indels: 0  
DB: 4 Gaps: 0

US-09-302-812-5 (1-3814) x US-09-511-477-6 (1-968)

QY 28 ATGAGTGGGCGCGCGCTGGAGCCCTGCACGAAAGCGCGTGGGGCGCGCTGGAAC 87  
Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20  
QY 88 TCTGCGCGACTGCTCGGACTCCCGGAGTTCCTCGCAGGAGGAGGCGGTCTCGAC 147  
Db 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeuAsp 40  
QY 148 CCCAAGGACGCTCCCGTCCAGTTCAGGTCCCTCCGTCCTCGCCAGCCTGCTCGGG 207  
Db 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
QY 208 CGGGCGGACCGCACAGAGGCAACGCCACCTCGTTTGTCTTCAACAAAGACTATTACT 267  
Db 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
QY 268 ACTTGGATGGTACTAAAGGACCCCAAGACAGCTGAATCAGAAAGTAAAGAAACAACAAT 327  
Db 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
QY 328 ACAAGAATTGACTCCATGATGAGTTCGTGAGAAAGATAACTTTTACCCACATAAGGTG 387  
Db 101 ThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPhelyrProHisLysVal 120  
QY 388 GAAAAATTGAAAAATGTTCTCAGCTAAATCTTGATAAATCACCCACAGAAAGAGTTCA 447  
Db 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer 140  
QY 448 CAGTATTGAACCAACAGCAGACTGCGAGTGTGTCAAGTGGCAGATGAAGGGAAGCAT 507  
Db 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160  
QY 508 GCAGAACAGCTTTTGGCAAGTGAAGCTCCCGCGGGGACTCCGCTACCCAAAGCAGCTTAGT 567  
Db 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
QY 568 AATGCTAACATTGGTCACTCACCCACACTGATGACCACTGACACAGATCATGAAGAA 627  
Db 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200  
QY 628 GACAGAGACATCAGCAGTTTCTTACCTATAAACTTGCAAAATCAAGCCACAGTA 687  
Db 201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220  
QY 688 GGAGATGGCAGGCGCAGAAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGAC 747  
Db 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240  
QY 748 TGTACAGGCTGTCAACAGGAGGAGGTGATGTGCTACAGAGAGTCTTGTGTGATGTT 807  
Db 241 CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal 260  
QY 808 GGTGCGGAGGACATTGGAACCTGGACCCCAAAAAATGACAAATTTGACTGGACAAGAAC 867  
Db 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlnGluSer 280  
QY 868 AGCCTAGGTGATTGCTCCTCCATTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTA 927  
Db 281 SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal 300  
QY 928 GACAACTCGAGAAACAGTTGTCAAGATTCAAGAGCAGATGAAGAAACAAGTCCAGTCTTT 987



Db 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
QY 988 GATGAGCAAGATGATCGTTCTCTCCAAACAGCAAAATAAATTTCAAGTTGCCAAGCAAGA 1047  
Db 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340  
QY 1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTGACTAAGGGAAGTGAAGTGTAGATTGCAT 1107  
Db 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
QY 1108 TTCCAATTGAAGGAGAAATAATGCTGGGACCAGTGAATTAATGCCAAGCCATCTGGA 1167  
Db 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAGTTCCAAGCAGCATGGAAGGATCTCT 1227  
Db 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
QY 1228 AAAATTACAGATCATTTTCATGAGAAATTTCCAAGTCAGAGGACAGAGAAAGAAACAATGT 1287  
Db 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420  
QY 1288 GAAGTCAGACATCAAGAAACAGAAAGGAAGATTCCAAAATACATCCCACTTAACCTCCCT 1347  
Db 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440  
QY 1348 CCAGAGAAGAAGTGGCTGGGAACCTCCTATTGAGGAAATGAGAAAAATGCTCGGTGGG 1407  
Db 441 ProGluLysLysTyrLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460  
QY 1408 ATCCATTTGCCCTTCTTAAGACCATCTGCAAGTCACACAGTGAATGCTGCGGTAGACCTT 1467  
Db 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
QY 1468 CTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTGTGGGATAAC 1527  
Db 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500  
QY 1528 AACATGTGAAAATGCTTGTTCGGAACAAAACCTTGACCTGTGGAAGATGAGAATGGT 1587  
Db 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
QY 1588 GAGCGAAGTGCAGGAGTAGGTGGGAGCTCATTCAGACTGCACCTTCTCAACAAATTCACA 1647  
Db 521 GluArgThrAlaGlySerArgTyrGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540  
QY 1648 CGACCCCAGAACTTGAAGGATGCGATTCTGGAATACAAATGTGGCATATTCTAAGAAATGG 1707  
Db 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560  
QY 1708 GACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGCCCAACAT 1767  
Db 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGlnHis 580  
QY 1768 TTATATCAGTCCATTTTACCTGACATGGTGAATAATGCACTCTGTCTGCCAAATATTGCG 1827  
Db 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCCAGCCAATACCACTCTGAAACAGAGAAGATGAATCATTTCTGTCACGATGTACAGGAA 1887  
Db 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCGCACATTTCCCGCAGCGAATGCCAAG 1947  
Db 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640  
QY 1948 ATGAAATCGGAGTATTCTAGTTACCCAGACATTAATCTCAATCGGTTGTTGAAGGACGT 2007  
Db 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAAGGAAACCCAGAAAAACTGAAAAACACTCTTCTGCTACTTTCCGAGAGTCACAGAG 2067  
Db 661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu 680

QY 2068 AAAAAACCTACAGGATTGGTGACATTTTACAAGACAGAGTCTTTGAAGATTTTCCAGAATGG 2127  
Db 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp 700  
QY 2128 GAAAGGTGTGAAAAAGCCTCTGACACGCTTACACGTCACCTACCTACGAGGGTACCATAGAAGGC 2187  
Db 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGGCCGAGGCATGCTACAGGTGGATTGCAAAATCGTTTGTGGAGGTGGTGTGACT 2247  
Db 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGACTTGTACAAGAAGAAATCAGATTTTAAATCAATCCTGAATTGTTTCA 2307  
Db 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTCACTGAGGTGCTGGATCAATGAGTGTCTTATTATCAGAGGTACTGAACAG 2367  
Db 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGCCCCGAGCCATGAAGATGGG 2427  
Db 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTyrAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAGAGCAGATTGGCAGCGGCTGTCGCTGAGAAAGTGAGACGCTTAACAAGGCT 2487  
Db 801 SerGluLysAspAspTyrGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACCTCGATCAGTTTGTGCTGAGAAAGTGAGACGCTTAACAAGGCT 2547  
Db 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840  
QY 2548 TACTGCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGA 2607  
Db 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGCTGTGGTGCCCTTTGGGGTGACGCTAGATTAAAAGCCCTTAATACAGATCCTG 2667  
Db 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
QY 2668 GCAGCTGCTGGGCTGAACGTGACGTGGTTTATTTCACTTTGGGAGTCAAGTTGATG 2727  
Db 881 AlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGAGACATTTACAGCATGCACACTTTCTTACCGAGAGGAAAGCTGGATTTGAAAAAGTG 2787  
Db 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
QY 2788 TACAAGTTATTGCTTAGATACATACTACAAATGAAGAATGCAGAACTGTCCACCCCTGGACCA 2847  
Db 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTATCCATTTCATATACCATGCTGTGTGAGTCAAGTGCAGAGACCCTGAC 2907  
Db 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
QY 2908 ATGCCAGGACAGAAAGGCAGGCACC 2931  
Db 961 MetProGlyGlnLysAlaGlyThr 968

RESULT 3

US-09-511-507-6  
; Sequence 6, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH



FILE REFERENCE: NIAD 201  
 CURRENT APPLICATION NUMBER: US/09/511,507  
 CURRENT FILING DATE: 2000-02-23  
 PRIOR APPLICATION NUMBER: 09/302,812  
 PRIOR FILING DATE: 1999-04-30  
 NUMBER OF SEQ ID NOS: 38

SEQ ID NO 6  
 LENGTH: 968  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 FEATURE:  
 US-09-511-507-6

Alignment Scores:

Pred. No.: 0 Length: 968  
 Score: 5164.00 Matches: 968  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 76.24% Indels: 0  
 DB: 4 Gaps: 0

US-09-302-812-5 (1-3814) x US-09-511-507-6 (1-968)

QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCAGAAAGCGCGTGGGGCGCGCTGGAAC 87  
 Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20  
 QY 88 TCTGCGCGACTGCTCGGACTCCCGAGCTTCCCTGGCAGGCAGAGCGTGTCTCGAC 147  
 Db 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeuAsp 40  
 QY 148 CCCAAGGACGCTCCCGTCCAGTTCCAGGGTCCCTCGCTCCGCGAGCTCGGCTCGGG 207  
 Db 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
 QY 208 CGGGCGGACCGCACAGAGGCAACGCCACCTCGTTGTTTCAAAACAAAGACTATTACT 267  
 Db 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
 QY 268 ACTTGGATGGATACTAAAGGACCCCAAGACAGCTGAATCAGAAAGTAAGAAACAAAT 327  
 Db 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
 QY 328 ACAAGAAATGATCCATGATGAGTTCTGTGCAAGAGATAAATTTTACCACATAAGGTG 387  
 Db 101 ThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyProHisLysVal 120  
 QY 388 GAAAAATGGAAATGTTCTCAGCTAAATCTTGATAAATCACCCACAGAAAGAGTTCA 447  
 Db 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer 140  
 QY 448 CAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGATGAAGGAGCAT 507  
 Db 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160  
 QY 508 GCAGAACAGCTTTTGGCAAGTGAGCTCCCGGGGACTCCGCTACCAAGCAGCTTAGT 567  
 Db 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
 QY 568 AATGCTAATGCTCAGTCACCCACACTGATGACACAGTGACACAGATCATGAAGAA 627  
 Db 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200  
 QY 628 GACAGAGACATCAGCAGTTCTTACACCTATAAAATTTGCAATACAAAGCCAAAGTA 687  
 Db 201 AspArgAspAsnGlnGlnPheLeuThrProLysLeuAlaAsnThrLysProThrVal 220  
 QY 688 GGAGATGGCCAGGACAGCAACTGTAGTGCAGTGGATCTCGCCAGTCTGTGAAGAC 747  
 Db 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240  
 QY 748 TGTACAGGCTGTCAACAGGAGGAGGTGGATGTCTACAGAGAGTCTTTGTTCAGATGTT 807

Db 241 CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal 260  
 QY 808 GGTCCCGAGGACATTTGAACTGGACCAAAAAATGACAAATGACTGGACAGAAAGC 867  
 Db 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280  
 QY 868 AGCCTAGGTGATTCGCTCCCTTCCATTGAGAAAGAAAGTGAGCTGAGTCAACATGATGTA 927  
 Db 281 SerLeuGlyAspSerProPheGluLysGluSerGluProGluSerProMetAspVal 300  
 QY 928 GACAACTCGAGAAACAGTTGTCAAGATTCAAGAGCAGATGAAGAAACAAAGTCCAGTCTTT 987  
 Db 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaaspGluGluThrSerProValPhe 320  
 QY 988 GATGAGCAAGATGATCGTTCTCCCAACAGCAAAATAAACTTTCAAGTTGCCAAGCAAGA 1047  
 Db 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340  
 QY 1048 GAAGCTGATGGCGATCTTAGGAAACCGTATTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107  
 Db 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
 QY 1108 TTCCAAATTTGAGAGAGAAATAATGCTGGGACCAGTGAATTAATGCCAAGCCATCTGGA 1167  
 Db 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
 QY 1168 AACTCTTCTAGCTTAATGTAGAGTGTAGAGTCCAAAGCAGCATGGAAAAAGGATTCT 1227  
 Db 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
 QY 1228 AAAATTACAGATCATTTTCATGAGAAATTTCCAAAGTCAGAGGACAGAAAGAAACAATGT 1287  
 Db 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420  
 QY 1288 GAAGTCAGACATCAAGAAACAGAAAGAGATTTCCAAATATACATCCCACTAACCTCCCT 1347  
 Db 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440  
 QY 1348 CCAGAGAGAGAGTGGCTGGAACTCCTATTGAGGAAATGAGAAAAATGCGCTCGGTGGG 1407  
 Db 441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460  
 QY 1408 ATCCATTTGCTTCTTAAGACCATCTGCAAGTCACACAGTGAATGTTGGGTAGACCTT 1467  
 Db 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
 QY 1468 CTGAGAGCAGCAGAGGTTCCGAAACCTTTCCAAACACATTACAAAGATTTGTGGATAAC 1527  
 Db 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500  
 QY 1528 AAACATGTGAAAAATGCCTTGTTCGGAACAAAACTTGTACCTGTGGAAGATGAGAATGGT 1587  
 Db 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
 QY 1588 GAGCGAATCGCAGGGAGTAGTGGGAGCTCATTCAGACTGCATTCCTCAACAAATTCACA 1647  
 Db 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540  
 QY 1648 CGACCCCAAGAACTTGAAGGATGCGATTCTGAAATACAAATGTGGCATATTCTAAGAAATGG 1707  
 Db 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560  
 QY 1708 GACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTTGAAGAGCAGAGCCCCAACAT 1767  
 Db 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluAlaGluAlaGlnHis 580  
 QY 1768 TTATATCAGTCCATTTTACCTGATGATGGTGAATTAATTCATCTGTCTGCCAAATATTTC 1827  
 Db 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
 QY 1828 ACCCAGCCAATACCACTCCCTGAACAGAAAGATGAATCATCTGTACAGATGTACAGGAA 1887  
 Db 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620



Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 616 GATCATGAAGAAGACAGACAGCAATCAGCAGTTTCTTACACCTATAAAACTTGCAATACA 675  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 676 AAGCCAAACAGTAGGAGATGGCGAGGCCAGA-----AGCAACTGTAAGTGCAAGTGA 726  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGCCAGTCTGTGAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGTACCA 786  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro 259  
QY 787 GAGAGTCTTTGTTCAGATGTTGTGTGCGCCAGGACATTGGAACTGGACCAAAATGACAAC 846  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 847 AAATTGACTGGACAAGACAGCCTAGGTGATTGCGCTCCATTGAGAAAGAAAGTGA 906  
Db 280 LysLeuIleArgGlnGlnSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
QY 907 CCGAGTCAACCAATCGATGTACACAACCTCGAGAAACAGCTTGTCAAGATTCAGAACGAGAT 966  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAAAT 1023  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
QY 1024 AAACCTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTGACT 1083  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1084 AAGGGAAGTGAAGTTAGATTGCATTTCCAATTGAA--GGAGAAATAATGCTGGGACC 1140  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAAT 1200  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAAGCAGCATGGAAAAGGGATTCTAAAATTACAGATCATTTTCATGAGATTTCCAAG 1260  
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAAAGAAAGCAATGTGAAGTCAGACATCAAAAGAACAGAAAGCAAGATT 1320  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1321 CCAAATACATCCCACTAACCTCCCTCCAGAGAGAAGTGGCTGGGAACCTCTATTGAG 1380  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1381 GAAATGAGAAAAATGCTCGGTGTGGGATCCATTGCTTCCCTTAAGACCACTCTGCAAGT 1440  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACACAGTGACTGTTGCGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1501 ACACATTACAAAGATTGTGGGATAACAAACATGTGAAAATGCCCTTGTTCGGAACAAAC 1560  
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1561 TTGTACCTGTGGAAGATGAGAATGGTGAGCGCACTGAGGAGTAGGTGGGAGCTCATT 1620  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1621 CAGACTGCACCTTCTCAACAAATTCACACGACCCCAAGAACTTGAAGGATGCGATTCTGAA 1680  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

QY 1681 TACAATGTGGCATATTCTAAGAAAATGGGACTTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740  
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCATGTCATTTTACCTGACATGGTGAAA 1800  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCACCTCTGTCTGCCAAATATTTGCACCCAGCCCAATACCACTCTCTGAAACAGAAATG 1860  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATTTCTGTACGATGTACAGGAACAGATCGCCAGTCTTTTAGCTTAATGCTTCTTC 1920  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TGCACATTTCCCGCAGCGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 1980  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 1981 AACTTCAATCGGTTGTTGAAGGACGTTTCATCAAGGAAAACAGAAAACCTGAAAACACTC 2040  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2041 TTCTGCTACTTTCCGAAGAGTCACAGAGAAAACCTACAGGATTGGTGACATTTTACAAGA 2100  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2101 CAGAGTCTTGAAGATTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACAC 2160  
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2161 GTCACTTACAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCA 2220  
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2221 AATCGTTTGTGGAGGTGGTGTGACTGGTGGCGGACTTGTACAAGAAAGAAATCAGATTT 2280  
Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
QY 2281 TTAATCAATCCTGAATGATTTTTCACGGCTGTTCACTGAGGTGCTGGATCACAAATGAG 2340  
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTACAGTGAATACACAGGCTATGCTGAACTTAT 2400  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2401 CGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACG 2460  
Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2461 GAGATCGTTGCCATTGACGCACCTTCACCTTCAGACGCTACCTCGATCAGTTTGTGCTTGA 2520  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCT 2580  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2581 GAAATCTTCTGTCAGTGGCCACCGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2641 AGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACCTGACGCTGTTTAT 2700  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 899  
QY 2701 TTCACCTTTGGGACTCAGAGTTGTAGAGACATTTACAGCATGCACACTTCTCTTACC 2760  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919



QY 2761 GAGAGGAAGCTGGATGTTGGAAAAAGTGTACAAGTTATTGCTTAGATACACTACATGAAGAA 2820  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 2821 TGCAGAAACTGTTCCACCCCTGGACACAGACATCAAGCTTTATCCATTTCATATACCATGCT 2880  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 2881 GTTGAGTCAAGTCAGAGACCACTGACATGCCAGGACAGAAGGCAGGCACC 2931  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976  
RESULT 5  
US-09-511-477-4  
; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-477-4

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 64.89% Indels: 10  
DB: 4 Gaps: 6

US-09-302-812-5 (1-3814) x US-09-511-477-4 (1-976)

QY 28 ATGAGTGGCGGCCCGCTGGAGCCCTGCACGAAGCG---CGCTGGGGCGCGCTGGA 84  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTyrGlyAlaAla--- 19  
QY 85 ACTTCTGCGCGGACTCCCTCGGACTCCCGAGCTTCCCTGGCAGGCAGAGCGCTGCTCTC 144  
Db 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39  
QY 145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCCTGCTCG 204  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGGCGGACCGCAGAGGCAACGCCACCTCGTTGTTTTCACAAACAAAGACTATT 264  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 265 ACTACTTGGATGGATACTAAAGGACCCCAAGACAGCTGAATCAGAA-----AGTAAA 315  
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 316 GAAACAACAATACAGAAATGACTCCATGATGAGTCTGTGCAGAAAGATAACTTTTAC 375  
Db 100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
QY 376 CCACATAAGGTGGAAAAATGTTCCCTCAGCTAAATCTTGATAAATCACCACACA 435

Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 436 GAAAAGAGTTCACAGTATTGAAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTyrGlnAsn 159  
QY 496 GAAGGGAAGCATGCAGAAACAGCTTTTGGCAAGTGAAGCTCCCGGGGGGACTCCGCTACCA 555  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 556 AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCTCCACCCACACTGATGACACAGTGACACA 615  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 616 GATCATGAAGAGACAGACAGACAATCAGCAGTTTCTTACACCTATAAACTTGAATAACA 675  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 676 AAGCCAAACAGTAGAGATGGCAGGCCAGA-----AGCAACTGTAGTCAGTGA 726  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGCCAGTCTGTGAAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGTACCA 786  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
QY 787 GAGAGTCCCTTTGTTCAGATGTTGGTCCCGAGGACATTTGGAACCTGGACCAAAAAATGACAAC 846  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 847 AAATTGACTGGCAAGAAAGCAGCCCTAGGTGATTCGCTCCCTTCATTTGAGAAAGAAAGTGA 906  
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
QY 907 CCTGAGTCACCAATGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAAGCAGAT 966  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAAAT 1023  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
QY 1024 AAATTTTCAAGTTGCCAAAGCAAGAGAGCTGATGGCGATCTTAGGAAACGGTATTTGACT 1083  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1084 AAGGGAAGTGAAGTTAGATTGCTATTTCCAAATTTGAA---GGAGAAATAATGCTGGGACC 1140  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAATGCCAAGCCATCTGGAAACTCTTCTAGCTTAATGTAGAGTGTAGAAGT 1200  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAAGCAGCATGGAAAAAGGGATTCTTAAATTTACAGATCATTTTCATGAGAAATTTCCAAG 1260  
Db 400 SerLysGlnHisGlyLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAAAGAAAGCAATGTGAAGTCAGACATCAAGAACAGAAAGGAAGATT 1320  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1321 CCAAAATACATCCACCTAACCTCCCTCCAGAGAAAGAGTGGCTGGAACTCCTATTGAG 1380  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1381 GAAATGAGAAAAATGCCTCGGTGCTGGGATCCATTTGCTTCCCTTAAAGACCATCTGCAAGT 1440  
Db 460 GluMetArgArgMetProArgCysGlyLeArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACACAGTGTGTTCCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499



QY 1501 ACACATTACAAAAGATTTGTGGGATAACCAACATGTGAAAATGCCCTTGTTCGGAACAAAAC 1560  
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisVallysMetProCysSerGluGlnAsn 519  
QY 1561 TTGTACCTTGTGGAAGATGAGAAATGGTCAGCGAACTGCAGGGAGTAGGTGGAGCTCATT 1620  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1621 CAGACTGCACTTCTCAACAAATTCACAGCAACCCAGAACTTGTAAGGATGCGATTCTGAAA 1680  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1681 TACAATGTGCAATATCTAAGAAATGGGACTTTACAGCTTTTGGTTGATTTCTGGGATAAG 1740  
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAA 1800  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCACTGTCTGCCAAATATTTGCACCCAGCCCAATACCACTCTCTGAAACAGAAGATG 1860  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATTTCTGTACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TGACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTANTCTAGTTACCCAGACATT 1980  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 1981 AACTTCAATCGTTTGTGGAAGAGCTTCATCAAGGAAACAGAAAAAACTGAAAAACATC 2040  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrIleu 679  
QY 2041 TTCTGCTACTTTTCGAAGAGTCCACAGAGAAAAACCTACAGGATTTGGTGACATTTACAGA 2100  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2101 CAGAGTCTTGAAGATTTTCCAGATGGGAAAGGTGTGAAAGCCCTCTGACACGCTTACAC 2160  
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2161 GTCACCTTACGAGGTACCATAGAAGCAACGCGCGAGGATGCTACAGGTGGATTTCGA 2220  
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2221 AATCGTTTTTGTGGAGTGTGTGACTGGTGGGACTTGTACAAAGAAGAAATCAGATT 2280  
Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
QY 2281 TTAATCAATCCTGAATTGATTGTTTACGGCTGTTCACTGAGGTGCTGATCAACAATGAG 2340  
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2341 TGCTTTATTATCACAGGTACTGAACAGTACAGTGAATACAGGCTATGCTGAACTTAT 2400  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2401 CGTTGGCCCCGAAGCCATGAAGATGGGAGTGAAAGGACGATTGGCAGCGCGCTGCACG 2460  
Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2461 GAGATCGTTGCCATTGACGCACTTCACTTCAGACGGCTACCTCGATCAGTTTGTGCTGAG 2520  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2521 AAAGTGAGACGTGAGCTTAAAGGCTTACTGCGGATTCCTCCGTCCCTGGAGTTCTCT 2580  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

QY 2581 GAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTTGGGGGTGACGCT 2640  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2641 AGATTAAAAAGCCTTAATACAGATCCTTGGCAGCTGCTGCGGCTGACCTGACGTGTTTAT 2700  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGATGCACACTTTCCTTACC 2760  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 2761 GAGAGAAAGCTGGATGTTTGGAAAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAA 2820  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTATATACCATGCT 2880  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAAAGGCAGGCACC 2931  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 6  
US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-507-4

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 64.89% Indels: 10  
DB: 4 Gaps: 6

US-09-302-812-5 (1-3814) x US-09-511-507-4 (1-976)

QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCACGAAAGCG---CGCTGGGGCGCGCTGGA 84  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19  
QY 85 ACTTCTGCGCGACTGCTCGGACTCCCGAGTTCCTGGAGGCTCCCTGGCAGGCGGTGTTCTC 144  
Db 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 145 GACCCCAAGGACGCTCCCGTCCAGTTCCAGGTCCCTCCGTCCCTCGCCAGCCTGCTCTC 204  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGGCGGACCGCACAGAGGCAACGCCACCTCGTTGTTTTCAAAACAAAGACTATT 264  
|||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||

```
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79
QY 265 ACTACTTGGATGATACATAAGGACCAAGACAGCTGAATCAGAA-----AGTAAA 315
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99
QY 316 GAAAAACAACAATACAAGAATTGACTCCATGATGATGATCTCTGTCGAGAAAGATAACTTTTAC 375
Db 100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119
QY 376 CCACATAAGGTGGAAAAATTGGAAAAATGTTCCCTCAGCTAAATCTTGATAAATCACCCACA 435
Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139
QY 436 GAAAGAGTTTACAGTATTTGAAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAAT 495
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159
QY 496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCTCCCGCGGGACTCCGCTACCA 555
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179
QY 556 AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCACACTGATGACACAGTGACACA 615
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199
QY 616 GATCATGAAGAAGACAGACACATCAGCAGTTTCTTACACCTATAAACTTGCAATACA 675
Db 200 AspSerGluGluAsnArgAspAsnGlnPheLeuThrValLysLeuAlaAsnAla 219
QY 676 AAGCCAAACAGTAGGAGATGGCAGGCCAGA-----AGCAACTGTAAGTGCAGTGA 726
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239
QY 727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGGTACCA 786
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259
QY 787 GAGAGTCTTTGTGAGATGTTGGTCCGAGGACATTGGAACCTGGACCAAAAAATGACAAC 846
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279
QY 847 AAATTGACTGGACAAGAACAGCCCTAGGTGATTCGCTCCATTTGAGAAAGAAAGTGAG 906
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299
QY 907 CCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGATTGTCAAGATTCAAGACAGAT 966
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319
QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAT 1023
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339
QY 1024 AAACCTTTCAAGTTGCCAAGCAAGAGAGCTGATGGCGATCTTAGGAAACGGTATTGACT 1083
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359
QY 1084 AAGGGAAGTGAAGTTAGATTGATTTCCAAATTTGAA---GGAGAAATAATGCTGGGACC 1140
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379
QY 1141 AGTGACTTAAATGCCAAGCCATCTGGAACTCTTTCAGCCTTAATGTAGACTGTAGAAGT 1200
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399
QY 1201 TCCAAGCAGCATGGAAGGATTTCTAAATTTACAGATCATTTTCATGAGAATTTCCAAG 1260
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419
QY 1261 TCAGAGGACAGAAAGAAACAATGTGAAGTCAGACATCAAGAACAGAAAGGAAGATT 1320
Db 420 AlaGluAspArgArgLysGluGlnTrpGlnThrLysHisGlnArgThrGluArgLysIle 439
```

---

```
QY 1321 CCAAAATACATCCCACCTAAACCTCCCTCCAGAGAAGAGTGGCTGGGAACCTCTATTGAG 1380
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459
QY 1381 GAAATGAGAAAAATGCCTCGGTGTGGATCCATTTCCCTTCCCTTAAGACCATCTGCAAGT 1440
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479
QY 1441 CACACAGTGACTGTTCCGGTAGACCTTCTGAGAGCAGGAGAGGTTCGGAACCTTTTCCA 1500
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499
QY 1501 ACACATTACAAAGATTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAAACAAAC 1560
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519
QY 1561 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGCAACTGCAGGGAGTAGGTGGGAGCTCAT 1620
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539
QY 1621 CAGACTGCACCTTCTCAACAAATTCACACGACCCCACTTGAAGGATGCGATTCTGAAA 1680
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559
QY 1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579
QY 1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATATCAGTCCATTTTACCTGACATGGTGA 1800
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599
QY 1801 ATTGCACCTCTGTGCCCCAAATATTGTCACCCAGCAATACCACCTCTCTGAAAACAGAA 1860
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619
QY 1861 AATCATTTCTGTACAGATGTACAGGAACAGATGCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920
Db 620 AsnHisSerIleThrMetSerGlnGlnGluIleAlaSerLeuAlaAsnAlaPhePhe 639
QY 1921 TGCACATTTCCCGACGGAATGCCAAGATGGAATCGGAGTATTCTAGTTACCCAGACATT 1980
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659
QY 1981 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAACCAAGAAACCTGAAAAACATC 2040
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679
QY 2041 TTCTGTACTTTTCAAGAGTCAACAGAGAAAAACCTACAGGATTGGTGACATTTTACAAGA 2100
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699
QY 2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAGGTGTGAAAAGCTCTGACACGCTTACAC 2160
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719
QY 2161 GTCACTTACAGGGTACCATAGAAAGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCA 2220
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739
QY 2221 AATCGTTTTTGTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAGAAGAAATCAGATT 2280
Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759
QY 2281 TTAATCAATCTGAATTGATTGTTTCAACGGCTGTTCACTGAGGTGGTGGATCACAATGAG 2340
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779
QY 2341 TGCTCTTATATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799
```

```
QY 2401 CGTTGGCCGAGCCATGAAGATGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACG 2460
Db |||||:::|||||
800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspCysGluArgArgCysThr 819
QY 2461 GAGATCGTTGCCATTGACGCACTTCACCTTCAGACGCTACCTCGATCAGTTTGTGCTGAG 2520
Db |||||:::|||||
820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839
QY 2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTCGGATTCTCCTCGCTGGAGTTCTTCT 2580
Db |||||:::|||||
840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859
QY 2581 GAAATCTTCTGCAGTGGCCACGGGAACTGGGGCTGTGGTGCCTTTGGGGTGACGCT 2640
Db |||||:::|||||
860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879
QY 2641 AGATTAAAGCCTTAATACAGATCTGGCAGCTGCTGCGGTGAACGTGACGTGGTTTAT 2700
Db |||||:::|||||
880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899
QY 2701 TTCACCTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760
Db |||||:::|||||
900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919
QY 2761 GAGAGGAAGCTGGATGTTGGAAGTGTACAAGTTATTGCTTAGATATACTACAATGAAGAA 2820
Db |||||:::|||||
920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939
QY 2821 TGCAAGAACTGTTCCACCCCTGGACCAAGCATCAAGCTTTATCCATTATATACCATGCT 2880
Db |||||:::|||||
940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959
QY 2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAGCAGGACGACCC 2931
Db |||||:::|||||
960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 7
US-09-302-812-2
; Sequence 2, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-302-812-2
Alignment Scores:
Pred. No.: 0 Length: 977
Score: 4332.50 Matches: 820
Percent Similarity: 89.86% Conservative: 57
Best Local Similarity: 84.02% Mismatches: 90
Query Match: 63.97% Indels: 9
DB: 4 Gaps: 5
US-09-302-812-5 (1-3814) x US-09-302-812-2 (1-977)
QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCACGAAA---GCCGCTGGGGCGCCGCTGGA 84
|||||:::|||||
```

```
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20
QY 85 ACTTCTGCGCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGAGCGCGTGTCTC 144
Db |||||:::|||||
21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40
QY 145 GACCCCAAGGACGCTCCCGTCCAGTTCCAGGTCCCTCCGTCCTCGCCAGCCTCGTCTCG 204
Db |||||:::|||||
41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60
QY 205 GGGCGGCGGACCGCACAGAGCAACGCCACCTCGTTTGTGTTTCAAAACAAAGACTATT 264
Db |||||:::|||||
61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80
QY 265 ACTACTTGGATGGATATACTAAAGCAACCCCAAGACAGCTGAATCAGAA-----AGTAAA 315
Db |||||:::|||||
81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100
QY 316 GAAAACAACAATACAAAGATTGACTCCATCATGATGATGTTCTGTGCAGAAAGATACTTTTAC 375
Db |||||:::|||||
101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120
QY 376 CCACATAAGGTGGAAAAATTGAAATGTTCTCTCAGCTAAATCTTGATAAATCACCCACA 435
Db |||||:::|||||
121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLysGlyPheAspLysSerProVal 140
QY 436 GAAAAGAGTTCACAGTATTTGAACCAACAGCAGAGCTGCGAGTGTGTGCAAGTGGCAGAT 495
Db |||||:::|||||
141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTyrGlnAsn 160
QY 496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAAGCTCCCGCGGGGACTCCGCTACCA 555
Db |||||:::|||||
161 GluGlyProHisSerGluArgLeuLeuGluSerGluProAlaValThrLeuValPro 180
QY 556 AAGCAGCTTAGTAACTAATGCTAATGTTGTTGTCAGTCAACCCACACTGATGACACAGTGCACA 615
Db |||||:::|||||
181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200
QY 616 GATCATGAAGAAGACAGACACAATCAGCAGTTTCTTACACCTATAAACTTGAATACA 675
Db |||||:::|||||
201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220
QY 676 AAGCCAACAGTAGGAGATGGCAG-----GCCAGAAAGCAACTGTAAAGTGCAGTGA 726
Db |||||:::|||||
221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240
QY 727 TCTCGCCAGTCTGTGAAAGACTGTACAGGTGTCAACAGGAGGAGGTGGATGTGCTACCA 786
Db |||||:::|||||
241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260
QY 787 GAGAGTCTTGTTCAGATGTTGTTGTCGCGAGACATTTGGAAGTGGACCAAAATGACAAC 846
Db |||||:::|||||
261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280
QY 847 AAATTGACTGGAACAAGAAAGCAGCTAGTGTGATTCCCTCCCTCCATTTGAGAAAGAAAGTGAG 906
Db |||||:::|||||
281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300
QY 907 CCTGAGTCACCAATGGATGTAGACACTGAGAAACAGTTGTCAAGATTCAGAAGCAGAT 966
Db |||||:::|||||
301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320
QY 967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTTCTCTCC---CAAACAGCAAT 1023
Db |||||:::|||||
321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340
QY 1024 AAACCTTTCAAGTTGCCAAGCAAGAGAGCTGATGGCGATCTTAGGAAACGGTATTGACT 1083
Db |||||:::|||||
341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360
QY 1084 AAGGGAAGTGAAGTTAGATTGATTTCATTTCCAATTTGAA---GGAGAAATATATGCTGGGACC 1140
Db |||||:::|||||
361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380
```







|  |         |   |     |  |  |  |  |  |  |
|--|---------|---|-----|--|--|--|--|--|--|
| Alignment Scores:                                  |         |   |     |  |  |  |  |  |  |
| Pred. No.:   | 0       | Length:   | 977 |  |  |  |  |  |  |
| Score:   | 4332.50 | Matches:  | 820 |  |  |  |  |  |  |
| Percent Similarity:                                | 89.86%  | Conservative:   | 57  |  |  |  |  |  |  |
| Best Local Similarity:                             | 84.02%  | Mismatches:   | 90  |  |  |  |  |  |  |
| Query Match:                                       | 63.97%  | Indels:   | 9   |  |  |  |  |  |  |
| DB:  | 4       | Gaps:   | 5   |  |  |  |  |  |  |
| US-09-302-812-5 (1-3814) x US-09-511-477-2 (1-977) |         |   |     |  |  |  |  |  |  |
| QY   | 28      | ATGAGTGGGGCCCGCTGGGAGCCCTGCACGAAA--GCGCGTGGGGCGCGCTGGA        | 84  |  |  |  |  |  |  |
| Db   | 1       | MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTyrAspAlaAla     | 20  |  |  |  |  |  |  |
| QY   | 85      | ACTTCTGCGCCGACTGCCTCGGACTCCCGAGCTTCCCTGGCAGGCAGAGCGTGTCTC     | 144 |  |  |  |  |  |  |
| Db   | 21      | ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgVal     | 40  |  |  |  |  |  |  |
| QY   | 145     | GACCCCAAGGACGCTCCCGTCCAGTTCAGGTTCCCTCCGTCCTCGCCAGCCTGCGTCTCG  | 204 |  |  |  |  |  |  |
| Db   | 41      | AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu  | 60  |  |  |  |  |  |  |
| QY   | 205     | GGCGGGCGGGACCCGACAGAGGCAACGCCACCTCGTTTGTTCCTTCAAAACAAAGACTATT | 264 |  |  |  |  |  |  |
| Db   | 61      | GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle  | 80  |  |  |  |  |  |  |
| QY   | 265     | ACTACTTGATGGATACTAAAGGACCCAGACAGCTGAATCAGAA-----AGTAAA        | 315 |  |  |  |  |  |  |
| Db   | 81      | ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys  | 100 |  |  |  |  |  |  |
| QY   | 316     | GAACAACAATAACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATACTTTTAC    | 375 |  |  |  |  |  |  |
| Db   | 101     | GluAsnAsnThrArgGluGluSerMetSerSerValGlnLysAspAsnPheTyr        | 120 |  |  |  |  |  |  |
| QY   | 376     | CCACATAAGGTGGAATAATTGGAAAATGTTCTCAGCTAAATCTTGATAAATCACCACA    | 435 |  |  |  |  |  |  |
| Db   | 121     | GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal  | 140 |  |  |  |  |  |  |
| QY   | 436     | GAAGAAGTTACAGTATTGTAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT    | 495 |  |  |  |  |  |  |
| Db   | 141     | GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn  | 160 |  |  |  |  |  |  |
| QY   | 496     | GAAGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGGGGAGCTCCGCTACCA    | 555 |  |  |  |  |  |  |
| Db   | 161     | GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro  | 180 |  |  |  |  |  |  |
| QY   | 556     | AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCACACTGATGACCAAGTGACACA    | 615 |  |  |  |  |  |  |
| Db   | 181     | GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr  | 200 |  |  |  |  |  |  |
| QY   | 616     | GATCATGAAGAAGACAGACAAATCAGCAGTTTCTTACACCTATAAACTTGCAAAATACA   | 675 |  |  |  |  |  |  |
| Db   | 201     | AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla  | 220 |  |  |  |  |  |  |
| QY   | 676     | AAGCCCAACAGTAGGAGATGGGCAG-----GCCAGAAGCAACTGTAAGTGCAGTGA      | 726 |  |  |  |  |  |  |
| Db   | 221     | LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys  | 240 |  |  |  |  |  |  |
| QY   | 727     | TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCA  | 786 |  |  |  |  |  |  |
| Db   | 241     | AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer     | 260 |  |  |  |  |  |  |
| QY   | 787     | GAGAGTCTTTGTTCAGATGTTGGTGGCCGAGGACATTGGAACTGGACCAAAAATGACAAC  | 846 |  |  |  |  |  |  |
| Db   | 261     | GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn  | 280 |  |  |  |  |  |  |
| QY   | 847     | AAATTGACTGGACAAGAAAGCAGCCTAGGTGATTTCGCCTCCATTTCAGAAAGAAAGTGAG | 906 |  |  |  |  |  |  |
| Db   | 281     | ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu  | 300 |  |  |  |  |  |  |
| QY   | 907     | CCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTTGTCAAGATTGAGAAGCAGAT | 966 |  |  |  |  |  |  |
| Db   | 301     | ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp  | 320 |  |  |  |  |  |  |

|    |      |   |      |
|----|------|---|------|
| QY | 967  | GAAGAAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCC---CAAAACAGCAAAAT | 1023 |
| Db | 321  | GluGluThrSerProGlyPheAspGluGlnGlnAspSerSerAlaGlnThrAlaAsn       | 340  |
| QY | 1024 | AAACTTTCAAGTTGCCAAGCAAGAGAGAGCTGATGGCGATCTTAGGAAACGGTATTGACT    | 1083 |
| Db | 341  | LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla    | 360  |
| QY | 1084 | AAGGGAAGTGAAGTGAAGTTCATTTCCCAATTTGAA---GGAGAAAATAATGCTGGGACC    | 1140 |
| Db | 361  | LysGlyGlyGluLeuArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet    | 380  |
| QY | 1141 | AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT    | 1200 |
| Db | 381  | AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn    | 400  |
| QY | 1201 | TCCAAGCAGCATGGAAAAAGGATTCTTAAATTAACAGATCATTTTCATGAGAATTTCCAAG   | 1260 |
| Db | 401  | SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys    | 420  |
| QY | 1261 | TCAGAGGACAGAAAGAAACAATGTGAAGTCAGACATCAAGAAACAGAAAGGAAGATT       | 1320 |
| Db | 421  | AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle    | 440  |
| QY | 1321 | CCAAAAATACATCCACCTAACCTCCCTCCAGAGAAAGTGGCTGGGAACTCTATTGAG       | 1380 |
| Db | 441  | ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu    | 460  |
| QY | 1381 | GAATGAGAAAAAATGCCTCGGTGGGATCCATTTGCCTTCCCTTAAGACCATCTGCAAGT     | 1440 |
| Db | 461  | GluMetArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn       | 480  |
| QY | 1441 | CACACAGTGAAGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA     | 1500 |
| Db | 481  | HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro    | 500  |
| QY | 1501 | ACACATTACAAAGATTGTGGGATAACAAACATGTGAAATGCCTTGTTCGGAACAAAAC      | 1560 |
| Db | 501  | ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn    | 520  |
| QY | 1561 | TTGTACCTCTGTGAAGATGAGAATGCTGAGCGAACTGCAGGGAGTAGTGGGAGCTCATT     | 1620 |
| Db | 521  | LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle    | 540  |
| QY | 1621 | CAGACTGCACTTCTCAACAATTCACACGACCCAGAACTTGAAGGATCGGATCTGAAA       | 1680 |
| Db | 541  | GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys    | 560  |
| QY | 1681 | TACATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTGTGATTCTGGGATAAG     | 1740 |
| Db | 561  | TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys    | 580  |
| QY | 1741 | GTACTTGAAGAAGCAGAGCGCCCAACATTATATCAGTCCATTTACCTGACATGGTGAAA     | 1800 |
| Db | 581  | ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys    | 600  |
| QY | 1801 | ATTGCACTGTCTGCCAAATATTTCACAGCAATACCTCCCTGAAACAGAGATG            | 1860 |
| Db | 601  | IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet    | 620  |
| QY | 1861 | AATCATTTCTGTACGATGTACAGAAATGTCAGCAAGATGAAATCGGAGTATTTAGCTTTC    | 1920 |
| Db | 621  | AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe    | 640  |
| QY | 1921 | TGCACATTTCCCGACGGAATGCAAGATGAAATCGGAGTATTTAGTATCCCGACATT        | 1980 |
| Db | 641  | CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle    | 660  |
| QY | 1981 | AACITTCATCGGTTGTTGAGGACGTTTCATCAAGGAAACAGAAACAACTGAAACACTC      | 2040 |
| Db | 661  | AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu    | 680  |

QY 2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCCTACAGGATTGGTGACATTACAAGA 2100  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAGGTGTGAAAAGCCCTCGACACGCTTACAC 2160  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
QY 2161 GTCACCTTACGAGGTACCATAGAAGCGCCGAGGCATGCTACAGGTGATTTTGCA 2220  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2221 AATCGTTTGTGGAGGTGGTGTGACTGGTGGGACTTGTACAAGAAATCAGATT 2280  
Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
QY 2281 TTAATCAATCCTGAATTGATTGTTTTCACGCTGTTTCACTGAGGTGTGATCACAATGAG 2340  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAACTTAT 2400  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2401 CGTTGGGCCGAAGCCATGAAGATGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACG 2460  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820  
QY 2461 GAGATCGTTGCCATTGACGCACCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCTGAG 2520  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCCTCGTCTGGAGTTCTTCT 2580  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2581 GAAAATCTTCTGCAGTGGCCACCGGAACTGGGCTGTGGTGCCTTTGGGGTGACGCT 2640  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
QY 2641 AGATTAAAGCCTTAATACAGATCCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTAT 2700  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2701 TTCACCTTTGGGGACTCAGAGTTGATGAGACATTTACAGCATGCACACTTTCCTTACC 2760  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 2761 GAGAGAACTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAA 2820  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
QY 2821 TGCAGAACTGTTCCACCCCTGGACCAAGATCAAGCTTTATCCATCATATACCATGCT 2880  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAGGCGAGGC 2928  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

RESULT 9  
US-09-511-507-2  
; Sequence 2, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-2

Alignment Scores:  
Pred. No.: 0 Length: 977  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 63.97% Indels: 9  
DB: 4 Gaps: 5

US-09-302-812-5 (1-3814) x US-09-511-507-2 (1-977)

QY 28 ATGAGTGGGGCCCGCCGGTGGAGCCCTGCACGAAA---GGCGCTGGGGCGCGCTGGA 84  
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
QY 85 ACTTCTGCGCGACTGCCCTCGGACTCCCGAGCTTCCCTGGCAGGAGCGGTCTC 144  
Db 21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
QY 145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGCTCCGCGCAGCCTGCTCTCG 204  
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
QY 205 GGGCGGCGGACCGCACAGAGCAACGCCACCTCGTTGTTTCAAAACAAAAGACTATT 264  
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
QY 265 ACTACTTGGATGGATACATAAGGACCCCAAGACAGCTGAATCAGAA-----AGTAAA 315  
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
QY 316 GAAACAAACAATACAGAATTGACTCCATGATGATGATTTCTGTGCAGAAAGATAACTTTTAC 375  
Db 101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120  
QY 376 CCACATAAGGTGGAAAAATTGAAAAATGTTCCCTCAGCTAAATCTTGATAAATCACCACCA 435  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 436 GAAAGAGATTACAGTATTGAAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495  
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCTCCCGCGGGGACTCCGCTACCA 555  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 556 AAGCAGCTTAGTAATGCTAACATTGGTTCAGTCAACCCACACTGATGACACAGTGACACA 615  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
QY 616 GATCATGAAGAAGACAGAGCAATCAGCAGTTTCTTACCTATAAAACTTGCAAAATACA 675  
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 676 AAGCCAAACAGTAGGAGATGGCAG-----GCCAGAGCAACTGTAAAGTGCAGTGA 726  
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGAGGAGGTGGATGTGTACCA 786  
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260

QY 1825 -----TGCACCCAGCCAATACCACTCCTGAAACAGAGATGAATCATTTCT 1869  
Db 119 AsnIleValSerGlyIleLysThrGlyLeuArgLeuLeuAsnSerGlnGlnAlaGlyIle 138  
QY 1870 GTCACGATGTCACAGAACAGATCGCCAGCTCTTTAGCTAATGCTTTCTTCGACATTT 1929  
Db 139 ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158  
QY 1930 CCCCAGCGGAATGCCAAGATGAAATCGGAGTATCTAGTTACCCAGACATTAATCAAT 1989  
Db 159 ProAspAsnArgGlyAlaLys-----HisLeuProValIleAsnPheAsp 174  
QY 1990 CGG-----TTGTTTGAAGGACGTTTCATCAAGGAACAGAAACCTGAAACACTCTTC 2043  
Db 175 HisLeuSerLeuTyrIleSerTyrSerGlnSerGlnGluSerLysIleArgCysIleMet 194  
QY 2044 TGCTACTTTTCGAAGAGTCACAGAGAAAAACCTACAGGATTTGGTGACATTTACAAAGACAG 2103  
Db 195 HisTyrPheGluArgPheCysSerCysValProIleGlyIleValSerPheGluArgLys 214  
QY 2104 -----AGTCTTGAAGATTTTCAGAATGGGAAGGTGTGAAAAGCCTCTGCACGCTTA 2157  
Db 215 IleThrAlaAlaProAspAlaAspPheTyrSerLysSerAspValSerLeu----- 231  
QY 2158 CACGTCACCTACGAGGGTACCATTAGAGGCAACGCGGAGGCATGCTACAGGTGGATTTT 2217  
Db 232 -----TyrGlnProAsp-----AsnAlaLeuGluValAspPhe 242  
QY 2218 GCAATCGTTTTGTTGGAGGTGGTGTGACTGGTGGGGGACTGTGTACAAAGAAATCAGA 2277  
Db 243 AlaAsnLysTyrLeuGlyGlyGlySerLeuSerArgGlyCysValGlnGluIleArg 262  
QY 2278 TTTTAAATCAATCTGAATGATTGTTTTCACGGCTGTTTCACTGAGGTGCTGGATCACAAT 2337  
Db 263 PheMetIleAsnProGluLeuIleAlaGlyMetLeuPheLeuProArgMetAspAsn 282  
QY 2338 GAGTGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACT 2397  
Db 283 GluAlaIleGluIleValGlyAlaGluArgPheSerCysTyrThrGlyTyrAlaSerSer 302  
QY 2398 TATCGTTGGCCGGAAGCCATGAAGATGGAGTGAAGAAAGGACGATTGGCAGCGGCTGC 2457  
Db 303 PheArgPheAlaGlyGluTyrIleAspLysLysAlaMetAspPropPheLysArgArg 322  
QY 2458 ACGGAGATCGTTGCCATTGACGCCTT-----CACTTCAGACGCTAC 2499  
Db 323 ThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLysAspIle 342  
QY 2500 CTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTC 2559  
Db 343 Cys-----LeuLeuArgGluIleAsnLysAlaLeuCysGlyPhe 355  
QY 2560 CTC-----CGTCCTGGA----- 2562  
Db 356 LeuAsnCysSerLysAlaTrpGluHisGlnAsnIlePheMetAspGluGlyAspAsnGlu 375  
QY 2563 -----CGTCCTGGA----- 2571  
Db 376 IleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrThrAlaSer 395  
QY 2572 -----GTTCTCTCTGAAAATCTT--- 2589  
Db 396 HisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnAsnLeuIle 415  
QY 2590 -----TCTGCAGTGGCCACG 2604  
Db 416 ArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspAspGlyValAlaThr 435  
QY 2605 GGAACCTGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAGCCTTAATACAGATC 2664  
Db 436 GlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIleGlnTrp 455

QY 2665 CTGGCAGCTGCTCGGCTGAACGTGACGTGGTT---TATTTACACCTTTGGGACTCAGAG 2721  
Db 456 LeuAlaAlaSerGlnThrArgArgPropPheIleSerTyrTyrThrPhe-GlyValGluAl 475  
QY 2722 TTGATGAGAGACATTTACAGCATGCACACTTCTTACCAGAGAGGAAAGCTGGATGTTGA 2781  
Db 475 a----- 475  
QY 2782 AAAGTGTACAACTTATTGCTTAGATACTACAATGAAGAATGCAGAAACTGTTCCACCC-- 2839  
Db 476 -----LeuArgAsnLeuAspGlnLeuProProTh 485  
QY 2840 -----CTGACCAGACATCAAGC-----TTTATCCATTTC 2868  
Db 485 rLysLeuSerSerArgLeuAspSerAlaAsnSerProSerLeuThrCysLeuMetHisSe 505  
QY 2869 ATATACCATGCTGTTGAGTCAAGTGCAGAGACCACCTGACATGCCAGGACAGAGCAGGC 2928  
Db 505 rHisThrValSerLeuHis----- 511  
QY 2929 ACCTGAGGAACAAGTGACTAGGACCTCTCTCAAAGAGACATCCTATTGAAATGTGGG 2988  
Db 512 -----PheProThrTrpAl 516  
QY 2989 TGTGATGCTGAATTGACTGAATCTGATCTAAGTGTGTATATAATCCACATTTGTAATCA 3048  
Db 516 aIleArgMetGluLeu--ArgAlaAspLeuArg----- 526  
QY 3049 AGGATGCAGTCTCTTCTGCATATGCAGTGTGTTTCTTGTTCATCTCTGTGGACATGCCTTT 3108  
Db 527 -----SerIleLeuGlnTyrLeuProLe 534  
QY 3109 AGACATGGCTTCTTCAATTTTCTTCTCTCTTCACTCTTTATTCTTTGATTTTTTTTCC 3168  
Db 534 uValAlaGlnSerSerSerLeuValTrpProProSerVal----- 547  
QY 3169 AACTTGATTTCTTGGGAAAACCTCAAGAAAGTTGCACTCAGCTTCTAGATCTTTCTCTTC 3228  
Db 548 -----GluGluGluLeuGlnThrIleSerArgGlyProSerGl 560  
QY 3229 CTGCTGTGTGTGTGTCCAGACTGTTTGGTGGCTAGCAGATACCATCAGACTTGGAGAA 3288  
Db 560 userMetValAsnSerGlyGluAlaLeu--AlaLeuHisIleThrAsnMetArgLysSe 579  
QY 3289 GTTACAAATCCAGAAATCTGAGTTTGTGCTGCA----- 3319  
Db 579 rLeuSerLeuAsnAlaSerAspLeuAlaProTyrAlaLeuGlnGlyTyrGlyLeuPhePh 599  
QY 3320 -GATTTACCTGTGAGCTTCTCA-----CTCCCAACCCCT 3351  
Db 599 eAspLysLysIleSerArgGluGluSerAlaAsnPhePheGlyGluValValProAlaLe 619  
QY 3352 TGTTAGGCTTGTGTGTCTACATTTTCAATTTTGGAA----- 3388  
Db 619 uCysArgLeuLeuLeuGlnLeuProSerMetLeuGluLysHisTyrGlnLysAlaAspHi 639  
QY 3388 ----- 3388  
Db 639 sValLeuAspGlyValLysSerGlyLeuArgLeuLeuGlyProGlnGluAlaGlyIleVa 659  
QY 3389 -----GTTGAAGTTTTTCTTATGTTACTTAAAGCTAGTATCTTTTAGGCT----- 3433  
Db 659 lLeuLeuSerGlnGluLeuIleAlaAlaLeuLeuAlaCysSerPhePheCysLeuPhePr 679  
QY 3434 -----AAAACTATTTCTATTAAAGGCAGACTAATTTCCAGTTTCTCTTTTGAAGA 3483  
Db 679 oGluValAspArgSerLeuLysAsnLeuGlnGlyIleAsnPheSerGlyLeuPhe----- 697  
QY 3484 CATCATCCCTATAAGTAACGGTTTTTTTTCGTCCTTTTTTCCCGAGCGCTATTTTAGAAGC 3543  
Db 698 -----SerPheProTyrMetArgHisCysTh 706  
QY 3544 TGGCCAAAGAGGAAAGA---AAATGTAGATAAAAGGATTTCTCTCGGATGCTATAAGAA 3600



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:16:17 ; Search time 55.3609 Seconds  
(without alignments)  
13253.917 Million cell updates/sec

Title: US-09-302-812-5  
Perfect score: 6773  
Sequence: 1 gggggactgtgtgctgcggg.....aatcattgtcagaaaaaaa 3814

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2\_1/USFTO\_spool\_p/6333148/runat\_26052004\_150054\_5689/app\_query.fasta\_1.12437  
-DB=PIR\_78 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148@cgn 1.1 298 @runat\_26052004\_150054\_5689 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: pir1.\*
- 2: pir2.\*
- 3: pir3.\*
- 4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 539.5 | 8.0         | 997    | 2 B84726 | probable poly(ADP- |
| 2          | 443.5 | 6.5         | 726    | 2 T21138 | hypothetical prote |
| 3          | 328   | 4.8         | 364    | 2 A84726 | probable poly(ADP- |
| 4          | 176   | 2.6         | 6642   | 2 T29757 | protein UNC-89 - C |
| 5          | 171   | 2.5         | 997    | 2 T43523 | cut17 protein - fi |
| 6          | 169   | 2.5         | 1165   | 2 T16420 | hypothetical prote |
| 7          | 166.5 | 2.5         | 699    | 2 I38073 | nucleolar phosphop |
| 8          | 166   | 2.5         | 1210   | 2 I39410 | AF-4 protein, spli |
| 9          | 165   | 2.4         | 1192   | 2 A71623 | probable secreted  |
| 10         | 163   | 2.4         | 532    | 2 T06029 | hypothetical prote |
| 11         | 163   | 2.4         | 2722   | 2 T20532 | hypothetical prote |
| 12         | 163   | 2.4         | 2738   | 2 E88320 | protein F07A11.6 f |
| 13         | 162   | 2.4         | 1213   | 2 A58198 | serine/proline-ric |
| 14         | 161.5 | 2.4         | 773    | 2 F90537 | lipoprotein [impor |

|    |       |     |      |          |                    |
|----|-------|-----|------|----------|--------------------|
| 15 | 160.5 | 2.4 | 1403 | 1 A47328 | natural killer cel |
| 16 | 159.5 | 2.4 | 990  | 2 I51618 | nucleolar phosphop |
| 17 | 159.5 | 2.4 | 1001 | 2 T16419 | hypothetical prote |
| 18 | 159.5 | 2.4 | 1115 | 2 T19137 | hypothetical prote |
| 19 | 157.5 | 2.3 | 559  | 2 B44265 | ENL (translocation |
| 20 | 157.5 | 2.3 | 1359 | 2 T34036 | hypothetical prote |
| 21 | 157.5 | 2.3 | 2211 | 1 KPBO5  | coagulation factor |
| 22 | 157.5 | 2.3 | 2526 | 2 T20531 | hypothetical prote |
| 23 | 156   | 2.3 | 1274 | 2 T37193 | enamelin matrix pr |
| 24 | 155.5 | 2.3 | 917  | 2 JC7799 | PARIS-1 protein -  |
| 25 | 155   | 2.3 | 1463 | 2 T30290 | AAS surface protei |
| 26 | 155   | 2.3 | 1489 | 2 S60416 | DNA helicase YGL15 |
| 27 | 155   | 2.3 | 1792 | 2 T08878 | supervillin P205 - |
| 28 | 155   | 2.3 | 2361 | 2 T25752 | hypothetical prote |
| 29 | 154.5 | 2.3 | 1560 | 2 T42727 | proliferation pote |
| 30 | 153   | 2.3 | 1611 | 2 T38236 | hypothetical prote |
| 31 | 153   | 2.3 | 5327 | 2 T13564 | microtubule-associ |
| 32 | 152.5 | 2.3 | 406  | 2 S38170 | SRP40 protein - ye |
| 33 | 152.5 | 2.3 | 1615 | 2 JC6510 | ras-responsive ele |
| 34 | 152   | 2.2 | 489  | 2 A45988 | dentin matrix acid |
| 35 | 151   | 2.2 | 1093 | 2 F88556 | protein B0464.5a f |
| 36 | 151   | 2.2 | 1320 | 2 JC5630 | TCOF1 protein - mo |
| 37 | 151   | 2.2 | 1791 | 2 T02345 | hypothetical prote |
| 38 | 150   | 2.2 | 1070 | 2 T30848 | Duffy receptor - p |
| 39 | 150   | 2.2 | 1337 | 2 T30291 | dextranase - Strep |
| 40 | 150   | 2.2 | 2910 | 2 T28156 | DNA-directed RNA p |
| 41 | 149.5 | 2.2 | 1125 | 2 E90598 | membrane nuclease, |
| 42 | 149.5 | 2.2 | 2128 | 2 I52577 | beta-spectrin - mo |
| 43 | 149.5 | 2.2 | 2137 | 1 SJHUB  | spectrin beta chai |
| 44 | 149   | 2.2 | 1401 | 2 T48079 | hypothetical prote |
| 45 | 149   | 2.2 | 1507 | 2 B47328 | natural killer cel |

ALIGNMENTS

RESULT 1

B84726

probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84726

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: B84726

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-997 <STO>

A;Cross-references: GB:AB002093; NID:G4887749; PIDN:AAD32285.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g31870

A;Map position: 2

Alignment Scores:

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 9.41e-30 | Length:       | 997 |
| Score:                 | 539.50   | Matches:      | 205 |
| Percent Similarity:    | 36.62%   | Conservative: | 88  |
| Best Local Similarity: | 25.62%   | Mismatches:   | 213 |
| Query Match:           | 7.97%    | Indels:       | 294 |
| DB:                    | 2        | Gaps:         | 27  |

US-09-302-812-5 (1-3814) x B84726 (1-997)

QY 1732 TGGGATAAGGTACTTGAAGAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGAC 1791

Db 79 PheAspGluLeuIleAspGluLysGluSerLysArgTrpPheAspGluIleIleProAla 98

QY 1792 ATGCTGAAATTCACCTCTGTCTGCCAATATT----- 1824

Db 99 LeuAlaSerLeuLeuLeuGlnPheProSerLeuLeuGluValHisPheGlnAsnAlaAsp 118



Job time : 248.584 secs

```
Db 785 sAlaGluAspLysGlnProProAlaGlyLysLysHisSerSerGluLysArgSerSerAs 805
QY 1837 -----ATACCACCTCTGAAACAGAAAGATG-----AATCA 1865
Db 805 pSerSerSerLysLeuAlaLysLysArgLysGlyGluAlaGluArgAspCysAspAsnLy 825
QY 1866 TTCTGTACAGATG-----TCACAGGAACAGATCGCCAGTCTTTTAGCTAA 1910
Db 825 sLysIleArgLeuGluLysGluIleLysSerSerSerSerSerHisLy 845
QY 1911 TGCTTTCTCTGCACATTTCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTA 1970
Db 845 sGluSerSerLysThrLysProSerArgProSerGlnSerSerLysLysGluMetLe 865
QY 1971 CCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAACCAAGAAAAA-- 2028
Db 865 uProProProValSer-----SerSerSerGlnLysProAlaLysPr 880
QY 2029 ----CTGAAAACACTCTTCTGTACTTTTCGAGAGTCACAGAG-----AAAAAAC 2075
Db 880 oAlaLeuLysArg-----SerArgArgGluAlaAspThrCysGlyGlnAspPr 896
QY 2076 TACAGGATTGGTGACATTTTACAAGACAGAGCTTGTGAAGAT-----TTTCCAGAATGGGA 2129
Db 896 oProLysSerAlaSerSerThrLysSerAsnHisLysAspSerSerIleProLysGlnAr 916
QY 2130 AAGGTGTGAAAGCCTCTGACACGCTTACACGTCACCTTACGAGGGGTACCATA----- 2181
Db 916 gArgValGluGlyLysGlySerArgSerSerGluHisLysGlySerSerGlyAspTh 936
QY 2182 -----GAAGGCAACGGCCGA---GGCATGCTACA 2207
Db 936 rAlaAsnProPheProValProSerLeuProAsnGlyAsnSerLysProGlyLysProGl 956
QY 2208 GGTGGATTTTGCAAAATCGTTTGTGGAGGTGGTGTGACTGGTGCGGACTT---GTACA 2264
Db 956 nValLysPheAspLysGln-----GlnAlaAspLeuHisMetAr 969
QY 2265 AGAAGAAATCAGATTTTAAATCAATCCTGAATGATTGTTTCACGGCTG----- 2313
Db 969 gGluAlaLysLysMetLysGlnLysAlaGluLeuMetThrAspArgValGlyLysAlaph 989
QY 2314 -----TTCACGTAGGTGCTGGATCACAATGAGTGTCTTATTATCAGAGTACTGAACA 2366
Db 989 eLysTyLeuGluAlaValLeuSerPheIleGluCysGlyIleAlaThrGluSerGluse 1009
QY 2367 GTACAGTGAA-----TACACAGGCTATGCTGAAACT-----TA 2399
Db 1009 rGlnSerSerLysSerAlaTySerValTySerGluThrValAspLeuIleLysPheIl 1029
QY 2400 TCGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAGGACGATTGGCAG----- 2448
Db 1029 eMetSerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlava 1049
QY 2449 -----CGGCGCTGCACGGAGATCGTTGCCATTGACGCACCTTCACCTCAGACGCTACCT 2501
Db 1049 lLeuCysMetArgCysGlnSerIle-----LeuAsnMetAlaMetPheArgCysLysLy 1067
QY 2502 CGATCAGTTTGTGCCTGAGAAAGTGAGACCTGAGCTTAAACAAGGTTTACTGCGGATTCTCT 2561
Db 1067 sAsp-----IleAlaIleLysTySerArgThrLeuAsnLysHisPhe----- 1081
QY 2562 CCGTCTGGAGTTCCTTCTGAAAATCTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGG 2621
Db 1082 -----GluSerSerSerLysValAlaGlnAlaProSerProCysIl 1095
QY 2622 TGCCTTTGGGGGT 2634
Db 1095 eAlaSerThrGly 1099
```

QY 15 TGCGGG-----TCCAGCATGAGTGC GGCGCCCGCTG-----GGA 50  
 Db 151 CysGlyProProAspSerGlnHisLeuThrGlnAspArgGlyGlnGluGlyPheGly 170  
 QY 51 GCCCTGCACGAAAGCGCTGGGGCGCCGCTGGAACCTTCGCGCCGACTGCCTCGGACTC 110  
 Db 171 SerSerHisHisLysLysGlyAspArgArg-AlaAspGlyAspHisCysAlaSerValTh 190  
 QY 111 CCGGAGCTTCCCTGGCAGGAGGCGTGTCTCTCGACCCCAAGGACGCTCCCGTCCAGTT 170  
 Db 190 rAspSerAlaProGluArgGluLeuSerProLeuIleSerLeuProSerPro----- 207  
 QY 171 CAGGGTCCCTCCCTCGCCAGCCTGCGTCTCGGGCGGGGGGACCGCACAGAGCAA 230  
 Db 208 ----ValProProLeuSerProIleHisSerAsnGlnGlnThrLeuProArgThrGlnG 226  
 QY 231 CGCCACCTCGTTTGTTCACAAACAAAGACTATTACTTGTGATGGATACTAAAGACC 290  
 Db 226 ySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyrCysProAlaLysSerPr 246  
 QY 291 CAAG-----ACAGCTGAATCAGAAAGTAAGAAACAACAATAACAAGAAATTGACTCCAT 344  
 Db 246 oLysAspLeuAlaValLysValHisAspLysGlu-----ThrProGlnAspSerLe 263  
 QY 345 GATGAGTTCTGTGCAGAAAGATAAC-----TTTTACCCACATAAGGT 386  
 Db 263 uValAlaProAlaGlnProProSerGlnThrPheProProProSerLeuProSerLysSe 283  
 QY 387 GGAATAATTGAAATGTTCTCTCAGCTAAATCTT-----GATAAATC 428  
 Db 283 rValAlaMetGlnGlnLysProThrAlaTyrValArgPrometAspGlyGlnAspGlnAl 303  
 QY 429 ACCCACAGAAAAG-----AGTTCACAGTATTGTAACCAACACAGCAGAC 470  
 Db 303 aProSerGluSerProGluLeuLysProLeuProGluAspTyr-----ArgGlnGlnTh 321  
 QY 471 T-----CGGAGTGTGTGCAAGTGGCAGAATGA 497  
 Db 321 rPheGluLysThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuLysMetPr 341  
 QY 498 AGGGAAGCATGCAGAAACAGCTTTTGGCAAGTGAAGCTCCCGGGGACTCCGCTACCAA 557  
 Db 341 oSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGluIleLeuLy 361  
 QY 558 GCAGCTTAGTAATGCTAAACATTGGTCAGTCAACCCACACTGATGACCACTGACACAGA 617  
 Db 361 sGluMetThrHisSer-----TrpProProLeuThrAlaIleHisThrProSerTh 379  
 QY 618 TCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACCTGCAAAATACAA 677  
 Db 379 rAlaGluProSerLys-----PheProPheProThrLysAspSerGlnHisVa 395  
 QY 678 GCCAACAGTAGGAGATGGGCAGGCCAGAGCAACTGTAAGTGCAGTGGATCTCGCCAGTC 737  
 Db 395 lSerSerValThrGlnAsnGlnLysGlnTyrAsp-----ThrSe 408  
 QY 738 TGTGAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACAGAGAGT---CC 794  
 Db 408 rSerLysThrHisSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspAspLeuGl 428  
 QY 795 TTTGTACAGATGTTGGTCCGAGGACATTGGAACCTGGACCAAAATAATGACAAACAAATTGAC 854  
 Db 428 nLeuSerAsp-----SerGluAsp-----SerAspSerGluGlnTh 440  
 QY 855 TGGACAAAGAACGAGCCTAGGTGATTTCGCTCCATTTTGAGAAAGAAAGTGAAGCTGAGTC 914  
 Db 440 rProGluLysProProSerSerSerAlaProProSerAlaProGlnSerLeuProGlu-- 459  
 QY 915 ACCAATGGATGTAGACAACACTCGAGAAACAGTTGTCAAGATTTCAGAACAGATGAAGAAAC 974  
 Db 460 -ProValAlaSerAlaHisSerSerSerAlaGluSerGluSerThrSerAspSerAspSe 479

QY 975 AAGTCCAGTCTTTGATGAGCAAGATGATCGTTCTCTCCAAACAGCAATAAACTTTCAAG 1034  
 Db 479 rSerSerAspSerGluSerGluSerSerSerSerSerSerGluGluAsnGluProLeuGl 499  
 QY 1035 TTGCCAAGCAAGAGAGCTGATGGCGATCTTAGGAAACGG-----TATTT 1079  
 Db 499 uThrProAlaProGluProGluProProThrThrAsnLysTrpGlnLeuAspAsnTrpLe 519  
 QY 1080 GACTAAGGGAAGTGAA-----GTTAG 1100  
 Db 519 uThrLysValSerGlnProAlaAlaProProGluGlyProArgSerThrGluProProAr 539  
 QY 1101 ATTGCATTTCCAAATTTGAAGGAGAAATAATATCTGGGACCACTTAATAATGCCAAG-- 1158  
 Db 539 gArgHisProGluSerLysGlySerSerAspSerAlaThrSerGlnGluHisSerGluSe 559  
 QY 1159 -----CCATCTGGAAACTCTTCTAGCCTT----- 1182  
 Db 559 rLysAspProProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHi 579  
 QY 1183 -----AATGTAGTGTAGAAAGTTCC----- 1203  
 Db 579 sProGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnTh 599  
 QY 1204 -----AAGCAG----- 1209  
 Db 599 rValGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgTh 619  
 QY 1210 -----CATGGAATAAGGGATTCTAAATTTACAGATCATTTTCATGAGAAATTTCCAAGTC 1262  
 Db 619 rSerLeuGlnGlyGluArgGluProGlyLeuLeuProTyrGlySerArgAspGlnThrSe 639  
 QY 1263 AGAGGACAGAGAAAGAAACAATGTGAAGTCAGACATCAAGAACAGAAAGAAAGATTC 1322  
 Db 639 rLysAspLysProLysValLysThrLysGlyArgProArgAlaAlaAlaSerAsnGluPr 659  
 QY 1323 AAAATACATCCCACTAACCTCCCTCCAGAGAGAAAGTGGCTGGGAACCTCTATTGAGGA 1382  
 Db 659 oLys-----ProAlaValProProSerSerGlu----- 668  
 QY 1383 AATGAGAAATAATGCCTCGGTGGGTATTCATTTGCCTTCTTAAGACCATCTGCAAGTCA 1442  
 Db 669 ----LysLysLysHisLysSerSerLeuProAlaProSer----- 680  
 QY 1443 CACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAC 1502  
 Db 681 -----LysAlaLeuSerGlyProGluProAlaLysAs 691  
 QY 1503 ACATTACAAAGATTGTGGGATAACAAACATGTGAAATGCCTTGTTCGGAACAAACTT 1562  
 Db 691 pAsnValGluAspArgThrProGluHisPheAlaLeuValProLeuThrGluSerGln-- 710  
 QY 1563 GTACCCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGTGGGAGCTCATTC 1622  
 Db 711 -GlyProProHisSerGlySerGlySerArgThrSerGlyCysArgGlnAlaValValva 730  
 QY 1623 GACTGCCTTCTCAACAAATTCACAGCACCCAGAACTTGAAGGATGCGATTCTGAAATA 1682  
 Db 730 lGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspThr----- 746  
 QY 1683 CAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTCTGGGATAAGGT 1742  
 Db 747 -----LysLe 748  
 QY 1743 ACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAT 1802  
 Db 748 uLeuSerProLeuArgAspThrProProProGlnSerLeu-----MetValLysil 765  
 QY 1803 TGCACTGTCTGCCAAATATTTGCACCCAGCCA----- 1836  
 Db 765 eThrLeuAspLeuLeuSerArgIleProGlnProProGlyLysGlySerArgGlnArgLy 785  
 QY 1836 ----- 1836









; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 71692

; LENGTH: 1297

; TYPE: PRT

; ORGANISM: Staphylococcus haemolyticus

US-10-282-122A-71692

## Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 0.000242 | Length:       | 1297 |
| Score:                 | 169.50   | Matches:      | 200  |
| Percent Similarity:    | 32.23%   | Conservative: | 149  |
| Best Local Similarity: | 18.47%   | Mismatches:   | 411  |
| Query Match:           | 2.50%    | Indels:       | 323  |
| DB:                    | 12       | Gaps:         | 47   |

US-09-302-812-5 (1-3814) x US-10-282-122A-71692 (1-1297)

```
QY 145 GACCCCAAGCAGCTCCCGTCCAGTTC-----AGGGTCCCTCCGTCC 186
Db 235 AspSerLysGluSerIleThrLysTyrSerSerThrValGluHisValProHisGlu 254
QY 187 TCGCCAGCTGCGTCTCGGGGCGGGGCGGACCGCACAGAGGCAACGCCACCTCGTTTGT 246
Db 255 SerAsnHisAlaIleAlaAspAsnSerAsnLysGluAspThrLysArgSerSerGlnLeu 274
QY 247 TTCAAACAAAGACTATTACTACTTGGATGGATATAAGGACCCCAAGACAGCTGAATCA 306
Db 275 AspSerSerIleThrIle-----GluAsnGluSerThrIleGluSer 288
QY 307 GAAAGTAAA---GAAACAAACAATAACAAGAATTGACTCCATGATG-----AGTTCT 354
Db 289 SerSerAsnThrSerAsnAsnAsnGluArgThrProAsnTyrSerLysArgAspAsnThr 308
QY 355 GTGCAGAAAGATAACTTTTACCACATAAAGGTG-----GAAAAATTG 396
Db 309 ValAsnIleGluAsnIleTyrAlaSerGlnIleValGluGluIleArgLysGluArgGlu 328
QY 397 GAAATGTTCTCAGCTAAATCTTGATAAATCACCACAGAAAGAGTTTCACAGTATTG 456
Db 329 ArgLysValGlnGlnLysArgLysPheLysGluAlaLeuGlnAsnLysArgGlnGlnThr 348
QY 457 AACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAG 516
Db 349 AspGluGluAspSerIleGlnArgAlaIleAspGluMetTyrAlaLysGlnAlaGluGln 368
QY 517 CTTTGTGGCA---AGTGAGCCTCCGCGGGGACTCCGCTACCAAGCAGCTTAGTAATGCT 573
Db 369 TyrThrGlyGluSerSerLeuAsnGlnGlyAspAsnValSerAsnLysSerAsnGluSer 388
QY 574 AACATGTGTCACTACCCCACT-----GATGACCCAC 606
Db 389 GluIleAspLysSerLysHisSerTyrHisSerLysAspLysSerLeuAlaAspGluHis 408
QY 607 -----AGTGACACAGATCATGAAGAAGACAGACAGACAATCAGCAGTTTCTT 651
Db 409 AsnArgLeuValGlnAsnGlnThrAspGluGlnThrAsnSerAspAsnValAspAsnGln 428
QY 652 ACACCTATAAACTTGCAAAATACAAAGCCCAACAGTAGGAGATGGCGAGCCAGAAC 711
Db 429 ThrGluValSerAsnGluSerLeuGluPro----- 438
QY 712 TGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGAG 771
Db 439 -----TyrAsnTyrGluGlu 443
QY 772 GTGGATGTGCTACCAGAGAGTCTTTGTGATGTTGTGTCGCGAGGACATT----- 822
Db 444 IleAspLeuAsnGlnValSerSerValGlnGlnValArgGlnAspAspValGlnValLys 463
QY 823 -----GGAACTGGACCAAAAAATGACAAACAATTTGACTGGACAGAAAGC 867
Db 464 AspValLeuGluGlnSerSerLysIleAsnAsnAsnLysValGluSerTyrSerAsn 483
```

```
QY 868 AGCCTAGGTGATTGGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGA 927
Db 484 GluLysPheAspAspTyrLeuGluAspThrAsnSerHisGluGluMetLeuHisAspAsp 503
QY 928 GACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAACAACAGTCCAGTCTTT 987
Db 504 AspLeuHisGluGlnValMetAspAspAspGluAsnGluGlyIleSerAsnLysThrThr 523
QY 988 GATGAGCAAGATGATCGTTTCCTCCCAACACAGCAAAT-----AAACTTTCAAGT 1035
Db 524 AspGluAsnAsnAspGluLysIleAspAspAlaAsnTyrArgGluIleAsnGluSerGlu 543
QY 1036 TGCCAAAGCAAGAGAAAGCTGATGGCGATCTTAGGAAACGGTATTTCACCTAAGGGAAGTAA 1095
Db 544 SerLeuMetGlnAspLysAlaAsnAspLeuLysPheAsnAspGluValAsnAsnSerGlu 563
QY 1096 GTTAGATTGCATTTCCAATTTGAAGGAGAAATAATAGCTGGGACCAGTGACTTAAATGCC 1155
Db 564 -----AsnGlnGlnAsnSerSerGluAsnAsnIleAsnAsnAlaValArgAsnAla 580
QY 1156 AAGCCATCTGGA-----AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
Db 581 ValSerSerAspIleGluTyrAlaThrAsnGluAspGluGluAsnAspGlu---ArgLeu 599
QY 1201 TCCAAAGCAGCATGGAAGGAAAGGATCTCTAAAT----- 1233
Db 600 AlaGlnAspThrAsnLysGluAspGlnLysLeuSerGlnSerGluAspIleGlnHisGlu 619
QY 1234 -----ACAGATCATTTTC 1245
Db 620 SerLeuAsnAsnGluAspValSerLeuThrSerAsnLysThrAspAspSerGluHisLeu 639
QY 1246 ATGAGAAATTTCCAAGTCAGAGCAGAGAGAAAGAAACAATGTGAAGTCAGACATCAAAGA 1305
Db 640 GluLysAspSerLeuAsnGluAspLysLysAlaGlu----- 651
QY 1306 ACAGAAAGGAAGATTCCAAATATACATCCCACCTAACCTCCCTCCAGAGAGAAGTGGCTG 1365
Db 652 -----ProSerPheAsnLysThrAsnLysAlaProGlnLysMetSerIle 666
QY 1366 GGA-----ACTCTATTGAGGAAATGAGAAAAATG 1395
Db 667 LysProGlySerLysProPheAsnValValMetThrProSerAspLysLysArgValMet 686
QY 1396 -----CCTCGGTGTGGGATCCATTTGCCTTCCCTTAAGACCATCT 1434
Db 687 AspAlaLysLysAsnSerValSerArgAsnLysValAsnValProGluLeuLysProGlu 706
QY 1435 GCAAGTCACACAGTACTGTTTCGGGTAGACCTTCTGAGAGCAGGAGGTTCCGAAACCT 1494
Db 707 ThrLysLysGluAla-----GlnAspGluLysMetAsnAlaGlu 719
QY 1495 TTTCCAACACATTACAAAGATTTGTGGATAACAACAT----- 1533
Db 720 PheAspAsnHisLeuAsnGluSerGlnLeuAsnSerAspGluSerSerAspPheAsnVal 739
QY 1534 GTGAAAATGCCCTTGTTCGGAACAAACTTGTACCCCTGTGAAGATGAG-----AAT 1584
Db 740 AlaSerLeuGluTyrAsnGluSerSerGluHisSerValGluLysAspAsnIleIleAsn 759
QY 1585 GGTGAGCGAACTCGAGGAGTAGGTGGGAGCTCATTTGAGTGCATCTCTC----- 1635
Db 760 AspGluAsnThrArgGluAsnGluHisGlnAspValAspAsnSerGlnAsnAspMet 779
QY 1636 -----AACAAATTCACACGACCCCGAAGCTTGAAGGATCGGATTCGAAATACAAT 1686
Db 780 ProLysGlyAsnGlnPheSerLysValGlnAsn-----SerAsn 792
QY 1687 GTGGCATATTTAAGAAAATGGGACTTTTACAGCTTTGGTT-----GATTTCTGGGATAAG 1740
Db 793 AsnGlnAsnAspAsnLysHisAspIleAsnGluPheValSerLysGluGlyTyrSerGlu 812
QY 1741 GTACTT-----GAAGAGCAGAGGCCCCAACATTTATATATCAGTCCATT 1782
```



QY 2302 GTTTCACGGCTGTTCACTGAGGTGCTGGATCAACAATGAGTGTCTTATTATATCAACAGGTACT 2361  
Db 216 -----TyrAsnGluArgIleGluIleValGlyVal 225  
QY 2362 GAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGCCCGAAGCCATGAA 2421  
Db 226 GluArgPheSerGlyTyrThrAspHisAlaSerSerPheArgPheSerArgAlaLysAla 245  
QY 2422 GATGGGAGTGAAGAGGACGATGGCAGCGGCTGCACGGAGATCGTTGCCATTGACGCA 2481  
Db 246 GluGlyArgGluGluAspProValGlyArgArgLysThrSerAspLeuser----- 262  
QY 2482 CTTCACTTCAGACGCTACCTCGATCAGTTTGTGCGCTGAGAAAAGTG 2526  
Db 263 -----GluLysTyrPheProProLysMet 270

RESULT 11  
US-10-115-482-18  
; Sequence 18, Application US/10115482  
; Publication No. US20030212257A1  
; GENERAL INFORMATION:  
; APPLICANT: Spytex, et al.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM  
; TITLE OF INVENTION: AND METHODS  
; TITLE OF INVENTION: OF USING THE SAME  
; FILE REFERENCE: 21404-322D  
; CURRENT APPLICATION NUMBER: US/10/115,482  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: 60/281,086  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/281,136  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/281,863  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/281,906  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/282,934  
; PRIOR FILING DATE: 2001-04-10  
; PRIOR APPLICATION NUMBER: 60/283,512  
; PRIOR FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/285,325  
; PRIOR FILING DATE: 2001-04-19  
; PRIOR APPLICATION NUMBER: 60/285,890  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 60/286,068  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/286,292  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/287,213  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: 60/288,257  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,134  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: 60/282,020  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/291,725  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/294,771  
; PRIOR FILING DATE: 2001-05-31  
; PRIOR APPLICATION NUMBER: 60/296,965  
; PRIOR FILING DATE: 2001-06-08  
; PRIOR APPLICATION NUMBER: 60/299,128  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 149  
; SEQ ID NO 18  
; LENGTH: 1489  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-115-482-18

Alignment Scores: 0.000162 Length: 1489  
Pred. No.:

Score: 172.00 Matches: 161  
Percent Similarity: 33.45% Conservative: 115  
Best Local Similarity: 19.52% Mismatches: 301  
Query Match: 2.54% Indels: 248  
DB: 15 Gaps: 39  
US-09-302-812-5 (1-3814) x US-10-115-482-18 (1-1489)  
QY 250 AAACAAAAGACTATTACTACTTGGATG-----GATACTAAAGGACCCAAAGACAGCT 300  
Db 418 LysGlnArgValLeuThrAsnTyrMetPheProGlnGlnProArgGlyAspGluAspPhe 437  
QY 301 GAATCAGAAAGTAAAGAAAACAACAATACAAAGATTGACTCCATGATGAGTTCTGTGCAG 360  
Db 438 GlnSerAspSerAspSerPheAsnProThrLeuTrpGluGlnArgGlnGlnArgMet 457  
QY 361 AAAGATAACTTT---TACCCACATAAGGTGGGAAAATTTGAAAATGTTCTCAGCTAAAT 417  
Db 458 ThrValAlaPheGluPheGluAspLysLysGluAspAspGluAsnAlaGlyLysValLys 477  
QY 418 CTTGAT---AAATCACCACACAGAAAAGAGTTACAGTATTGAAACCAACAGCAG- 468  
Db 478 LeuSerCysGlnAlaProTrpGluArgGlyGlnArgGlyIleThrLeuGlnProAlaArg 497  
QY 469 ---ACTGCGAGTGTGTC---AAGTGGCAGAAATGAAGGAAGCATGCAGAACAGCTT--- 519  
Db 498 LeuSerGlyAspCysCysThrProTrpAlaArgCysAspGlnGlnIleGlnAspMetPro 517  
QY 520 -----TTGGCAAGTGAGCCT 534  
Db 518 ValProGlnAsnAspProGlnLeuAlaTrpGlyCysIleSerGlyLeuGlnGlnGluArg 537  
QY 535 CCCGCGGGGACTCCGCTACCAAGCAGCTTAGTAATGCT----- 573  
Db 538 SerMetCysThrProLeuProValAlaAlaGlnSerThrThrLeuProSerLeuSerGly 557  
QY 574 -----AACATTGGTCAGTCACTACCCACACACTGATGACACAGTACACAGAT 618  
Db 558 ArgGlnValGluIleAsnLeuLysArgTyrPro-----ThrPro 570  
QY 619 CATGAAGAAGACAGACAGACAATCAGCAGTTTCTTACACCTATAAACTTGCATAAAG 678  
Db 571 TyrProGluAspLeuLysAsn-----MetValLysSerValGlnAsnLeuValGlyLys 588  
QY 679 CCAACAGTAGGAGATGGCGAGGCCAGAACAACTGTAAGTGCAGTGGATCTGCCAGTCT 738  
Db 589 ProSerHisGlyValArgValGluAsnSerAsn---ProThrAlaAsnThrGluGlnThr 607  
QY 739 GTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTCTTACAGAGAGTCTTTG 798  
Db 608 ValLysGlu-----LysTyrGluHisLysTrpProValAlaProLys----- 621  
QY 799 TCAGATGTTGGTCCCGAGGACATTGGAACTGGACCAAAAATAATGACAAATGACTGGA 858  
Db 622 ---GluIleThrValGluAspSerPheValHisProAlaAsnGluMetArgIleGlyGlu 640  
QY 859 CAAGAAAGCAGCCTAGGTGATTGCTCCATT----- 891  
Db 641 LeuHisProSerLeuAlaGluThrProLeuTyrProProLysLeuValLeuLysLys 660  
QY 892 ---GAGAAAGAAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGT 948  
Db 661 AspLysLysGluSerThrAspGluSerGluValAspLysThrHisCysLeuAsnAsnSer 680  
QY 949 CAAGATTTCAGAACGACAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCC 1008  
Db 681 ValSerSerGlyThrTyrSerAspTyrSerPro-----SerGlnAlaSerSerGlySer 698  
QY 1009 TCCCAACACAGCAATAAATCAAGTTGCCAAGCA----- 1044  
Db 699 SerAsnThrArgValLysValGlySerLeuGlnThrThrAlaLysAspAlaValHisAsn 718  
QY 1044 ----- 1044





Db 1589 SerProThrLysLysSerGluAsnGluValLysSerProThrLysLysLysGluLysSerPro 1608  
QY 553 -----CCAAAGCAGCTTAGTAATGCTAAACATTGGT 582  
Db 1609 GluLysSerValValGluGluLeuLysSerProLysGluLysSerProGluLysAlaAsp 1628  
QY 583 CAGTCACCCACACT-----GAT 600  
Db 1629 AspLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaThrGluAsp 1648  
QY 601 GACCACAGTGACACAGATCATGAAGAA-----GACAGAGACAATCAGCAGTTCTTACA 654  
Db 1649 ValLysSerProThrLysLysGluLysSerProGluLysValGluGluLysProThrSer 1668  
QY 655 CCTATAAACTTGCAAATACAAAGCCAAACAGTAGGAGATGGGCAGGCCAGAACCACTGT 714  
Db 1669 ProThrLysLysGluSerSerProThrLysLysThrAspAspGluValLysSerProThr 1688  
QY 715 AAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTG 774  
Db 1689 LysLysGluLysSerProGlnThrValGlu----- 1698  
QY 775 GATGTGCTACCAGAGATCTTTGTGCAGATGTTGGTGCAGGACATTTGGAACCTGGACCA 834  
Db 1699 -----GluLysProAlaSerProThrLysLysGlu----- 1708  
QY 835 AAAAAATGACAAACAAATTGACTGGACAAGAA---AGCAGCTAGGTGATTCGCCTCATTT 891  
Db 1709 LysSerProGluLysSerValValGluGluValLysSerProLysGluLysSerProGlu 1728  
QY 892 GAGAAAGAAAGTGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAA 951  
Db 1729 LysAlaGluGluLysProLysSerPro-----ThrLysLysGluLysSerProGlu 1745  
QY 952 GATTGAGAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCC 1011  
Db 1746 LysSerAlaAlaGluGluValLysSerProThrLysLysGluLysSerProGluLysSer 1765  
QY 1012 -----CAACAGCAAAATAAACTTTCAAGTTGCCAAGCAAGCAAGAGCTGAT 1056  
Db 1766 AlaGluGluLysProLysSerProThrLysLysGluSerSerProValLysMetAlaAsp 1785  
QY 1057 GCGCATCTTAGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCAATTTCCAATTT 1116  
Db 1786 AspGluValLysSerProThrLysLysGluLysSerProGluLysValGlu----- 1802  
QY 1117 GAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCT 1176  
Db 1803 -----GluLysProAlaSerProThrLysLysGluLysThrProGluLysSerAlaAla 1820  
QY 1177 AGCCTTAATCTAGAGTGTAGAAGTTCCAAAGCAGCATGGAAGAAAGGATTCTAAATAATACA 1236  
Db 1821 Glu-----GluLeuLysSerProThrLysLysGluLysSerProSerSerProThr 1837  
QY 1237 GATCATTTCATGAGAATTTCCAAAGTCAGAGCAGACAGAAAGAAACAATGT---GAAGTC 1293  
Db 1838 -----LysLysThrGlyAspGluSerLysGluLysSerProGluLys 1851  
QY 1294 AGACATCAAGAACAGAAAGGAAGATTCCAAAATACATCCCACTAACCTCCCTCCAGAG 1353  
Db 1852 ProGluGluLysProLysSerProThrProLysLysSerProProGlySerProLysLys 1871  
QY 1354 AAGAAGTGGCTGGAACTCCTATTGAGGAATAGAGAAAATGCCTCGGTGTGGATCCAT 1413  
Db 1872 LysLys-----SerLysSerProGluAlaGluLysPro----- 1883  
QY 1414 TTGCCTTCCTTAAGACCATTCTGCAAGTCACACAGTGACTGTTCCGGGTAGACCTTCTGAGA 1473  
Db 1884 AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAlaHis 1903  
QY 1474 GCAGGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAGATTTGTGGATAACAACAT 1533  
Db 1904 Phe---GluValValGluHisAlaThrGluCysLysLysTrpPheLeuAspGlyLysGlu 1922

QY 1534 GTGAAAAATGCCTTGTTCGGAACAAAACTTGTACCCCTGTGGAAGATGAGAATGGTGAGCGA 1593  
Db 1923 Ile-----ThrThrAlaGlnGlyValThrValSerLysAspGlnPheGluPhe 1939  
QY 1594 ACTGCAGGGAGTAGGTGGAGCTCATTGAGACTGCTCTCAACAAATTCACACGACCC 1653  
Db 1940 ArgCysSer-----IleAspThrThrMetPheGlySerGlyThr----- 1952  
QY 1654 CAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCT----- 1698  
Db 1953 -----ValSerValValAlaSerAsnAlaAlaGlySerValGluThrLysThr 1968  
QY 1699 -----AAGAAATGGGACTTTTACA----- 1716  
Db 1969 GluLeuLysValLeuGluThrProLysGluThrLysLysProGluPheThrAspLysLeu 1988  
QY 1717 -----GCTTTGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACAT--- 1767  
Db 1989 ArgAspMetGluValThrLysGlyAspThrValGlnMetAspValIleAlaLeuHisSer 2008  
QY 1768 ---TTATATCAG-----TTCCTCTCTGCACA-----TTCCTCCCGA-----TCC 1779  
Db 2009 ProLeuTyrLysTrpTyrGlnAsnGlyAsnLeuLeuGluAspGlyLysAsnGlyValThr 2028  
QY 1780 ATTTTACCTGACATGGTGAATAATTGCACCTCTGTCTGCAAAATATTTGCACCCAGCCAATA 1839  
Db 2029 IleLysAsnGluGluAsnLysSerSerLeuIleProAsn-----AlaGlnAspSer 2046  
QY 1840 CCACCTCTGAAACAGAAAGATGAATCATTTCTGACGATGTACAGGAACAGATCGCCAGT 1899  
Db 2047 GlyLysIleThrValGluAlaSerAsnGluValGly---SerSerGluSerSerAlaGln 2065  
QY 1900 CTTTGTAGCTAATGCTTTCTTCTGCACA-----TTCCTCCCGA----- 1935  
Db 2066 LeuThrValAsnProProSerThrThrProIleValValAspGlyProLysSerValThr 2085  
QY 1936 -----CGAATGCCAAGATGAATCGAGTATTCTAGTTACCCA----- 1974  
Db 2086 IleLysGluThrGluThrAlaGluPheLysAlaThrIleSerGlyPheProAlaProThr 2105  
QY 1975 -----GACATTAACCTTCAATCGTTTGTGTAAGGACGTTTCATCAAGAAACACAGAA 2025  
Db 2106 ValLysTrpThrIleAsn---GluLysIleValGlu-----GluSerArgThrIleThr 2122  
QY 2026 AAACCTGAAACACTCTCTGCTACTTTTCGAAGAGTACACAGAGAAAAA-----CCTACA 2079  
Db 2123 ThrIleLysThrGluAspValTyrThrLeuLysIleSerAsnAlaLysIleGluGlnThr 2142  
QY 2080 GGATTGGTGACATTTACAAGACAGAGT-----CTTGAAGATTTT 2118  
Db 2143 GlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAspSerLysGlnAlaAspLeu 2162  
QY 2119 CCAGAAATGGGAAAGGTGTGAAAGCCTCTGACAGGCTTACACGCTCCTACGAGGTACC 2178  
Db 2163 LysValGluProAsnValLysAlaProLysPheLysSerGlnLeuThrAspLysValAla 2182  
QY 2179 ATAGAAGGCCAACGGC---CGAGGCTGCTACAGGTGGATTTTGCATAATCGTTTGTGGA 2235  
Db 2183 AspGluGlyGluProLeuArgTrpAsnLeuGluLeuAsp----- 2195  
QY 2236 GGTGGTGTGACTGGTGGGACTTTGTACAAGAAATAATCAGATTTTAAATCAATCCTGAA 2295  
Db 2196 GlyProSerProGlyThr-----GluValSerTrpLeuLeuAsnGlyGln 2210  
QY 2296 TTGATTGTTTACGGCTGTTCACTGAGGTGCTGGATCAACAATGAG---TGTCTTATTATC 2352  
Db 2211 ProLeuThrLysSerAspThrValGlnValValAspHisGlyAspGlyThrTyrHisVal 2230  
QY 2353 ACAGGTACTGAACAGTACAGTGAATACACAGGC 2385  
Db 2231 ThrIleAlaGluAlaLysProGluMetSerGly 2241



|    |      |  |      |
|----|------|--|------|
| Db | 108  | GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal   | 127  |
| QY | 904  | GAGCTGACTCACCAATGGATGTA  | 927  |
| Db | 128  | AsnAsnLysSerAsnIleAspValAlaIleAsnSerAspGluAspAspGluLeuValLeu   | 147  |
| QY | 928  | ---GACAACTCGAGAAAACAGTTGTCAAGATTGAGAACGAGATGAAGAAAACAAGTCCAGTC | 984  |
| Db | 148  | GluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnGlnLeuSerGlnAspLeu   | 167  |
| QY | 985  | TTT---GATGAGCAAGAT   | 999  |
| Db | 168  | PheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrThrGln   | 187  |
| QY | 1000 | -----GATCGTTCCTCCCAAACAGCAAATAAATCTTCAAGTTGCCAAGCAAGA          | 1047 |
| Db | 188  | LeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThr   | 207  |
| QY | 1048 | GAAGCTGATGGCGATCTTAGGAAACGGTATTTTGACTAAGGGAAGTGAAGTTAGATTGCAT  | 1107 |
| Db | 208  | GluAlaAspSer   | 211  |
| QY | 1108 | TTCCAAATTTGAAGGAGAAAATAATGCTGGACCACTGACTTAAATGCCAAGCCATCTGGA   | 1167 |
| Db | 212  | ---ThrPheValGlyGluAspSerLysAlaThrLysThrValArgThrSerSerSerSer   | 230  |
| QY | 1168 | AACTCTTAGCCTTAATGTAGAGTGTAGAAGTTCCAAAGCATGGAAAAAGG             | 1221 |
| Db | 231  | PheLeuSerThrValSerThr---CysGluAlaProAlaLysGlyArgAlaArgMetTyr   | 249  |
| QY | 1222 | GATTCTAAATATACAGATCATTTTCATGAGAATTTCCAAATCAGAG                 | 1266 |
| Db | 250  | GlnLysGluLeuGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnPro   | 269  |
| QY | 1267 | -----GACAGAAGAAAAGAAACAATGTGAAGTCAGATCAGACATCAAGA              | 1305 |
| Db | 270  | AspLeuAsnLysValAspProAspArgAsnTyrArgTyrCysThrIle               | 285  |
| QY | 1306 | ACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTG   | 1365 |
| Db | 286  | -----ProAsnPheProAlaSerGln                                     | 292  |
| QY | 1366 | GGAACTCCTATTGAGGAATGAGAAAATGCCTCGTGTGGATCCATTGCTTCCTTA         | 1425 |
| Db | 293  | GlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeu                  | 307  |
| QY | 1426 | AGACCATCTGCAAGTCACACAGTGACTGTTCCGGTAGACCTTCTGAGAGCAGGAGAGTT    | 1485 |
| Db | 307  | -----  | 307  |
| QY | 1486 | CCGAAACCTTTTCCAACACATTACAAAGATTTCTGGGATAACAACAATGTGAAATGCCT    | 1545 |
| Db | 308  | -----ProGlnArgTyrArgGluPhe                                     | 314  |
| QY | 1546 | TGTTCGGAACAAAACCTTGTAACCTGTGGAAGATGAGAATGGTGAGCGA              | 1593 |
| Db | 315  | -----AspSerArgGlyArgArgAspSerTyr                               | 324  |
| QY | 1594 | -----ACTGCAGGG   | 1602 |
| Db | 325  | PheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPhe   | 344  |
| QY | 1603 | -----AGTAGGTGGAGACTCATTCAGACTGCACCTTCTCAACAAA                  | 1641 |
| Db | 345  | MetPheValGlyLeuLeuHisAsnMetTrpGluPhe                           | 356  |
| QY | 1642 | TTCACACGACCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCA             | 1692 |
| Db | 357  | -----AspProAspIleThrTyrLysLeuProAlaLeuGlu                      | 368  |
| QY | 1693 | -----TATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTCTGGGATAAGGTACTT     | 1746 |

|      |     |     |     |     |     |     |     |       |     |     |     |     |     |     |     |     |     |     |      |      |
|------|-----|-----|-----|-----|-----|-----|-----|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|
| 369  | Met | Tyr | Tyr | Lys | Glu | Met | Ser | Glu   | Leu | Val | Gly | Arg | Glu | Val | Leu | Glu | Lys | Phe | Ala  | 388  |
| 1747 | GA  | AG  | AG  | CAG | AGG | CC  | CA  | CA    | T   | T   | T   | A   | T   | A   | T   | A   | T   | A   | T    | 1806 |
| 389  | Arg | Val | Ala | Arg | Ile | Ala | Lys | Thr   | Ala | Glu | Asp | Ile | Leu | Pro | Glu | --- | Arg | Ile | Tyr  | 406  |
| 1807 | CT  | CT  | GT  | CT  | GC  | CA  | A   | A     | T   | A   | T   | T   | T   | G   | C   | A   | C   | C   | C    | 1866 |
| 407  | Arg | Leu | Val | Gly | Asp | Val | --- | ---   | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 413  |      |
| 1867 | T   | C   | T   | G   | T   | C   | A   | G     | A   | T   | G   | T   | C   | A   | G   | A   | A   | C   | A    | 1926 |
| 414  | Ser | Ala | Thr | Leu | Ser | His | Lys | Gln   | Cys | Ala | Ala | Leu | Val | Ala | Arg | Met | Phe | Phe | 431  |      |
| 1927 | T   | T   | T   | C   | C   | C   | G   | A     | G   | A   | A   | T   | G   | C   | C   | A   | A   | T   | G    | 1986 |
| 432  | --- | --- | --- | --- | --- | --- | --- | ---   | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 440  |      |
| 1987 | A   | A   | T   | C   | G   | T   | T   | G     | A   | G   | G   | A   | C   | T   | T   | C   | A   | A   | A    | 2046 |
| 441  | Cys | Arg | Ile | Leu | Ser | Ser | Asp | Lys   | Ser | Ile | Cys | Val | Glu | Lys | Leu | Lys | Phe | Leu | 460  |      |
| 2047 | T   | A   | C   | T   | T   | C   | G   | A     | G   | A   | A   | A   | A   | A   | A   | A   | A   | A   | A    | 2100 |
| 461  | Tyr | Phe | Asp | Lys | Met | Ser | Met | Asp   | Pro | Pro | Asp | Gly | Ala | Val | Ser | Phe | Arg | Leu | 480  |      |
| 2101 | C   | A   | G   | A   | G   | T   | T   | G     | A   | A   | G   | A   | T   | T   | T   | --- | --- | --- | 2154 |      |
| 481  | Met | Asp | Lys | Asp | Thr | Phe | Asn | Glu   | Glu | Trp | Lys | Asp | Lys | Lys | Leu | Arg | Ser | Leu | 500  |      |
| 2155 | T   | T   | A   | C   | A   | G   | T   | C     | A   | T   | T   | A   | G   | G   | T   | A   | C   | A   | G    | 500  |
| 501  | Val | Glu | Phe | Phe | Asp | Glu | Met | Leu   | Ile | Glu | Asp | Thr | Ala | Leu | --- | Cys | Thr | Gln | 519  |      |
| 2215 | T   | T   | T   | C   | A   | A   | A   | T     | C   | G   | T   | T   | G   | G   | A   | G   | T   | G   | T    | 2274 |
| 520  | Phe | Ala | Asn | Glu | His | Leu | Gly | Gly   | Val | Leu | Asn | His | Gly | Ser | Val | Gln | Glu | Ile | 539  |      |
| 2275 | A   | G   | A   | T   | T   | T   | A   | A     | T   | C   | A   | T   | T   | G   | A   | T   | T   | G   | A    | 2334 |
| 540  | Arg | Phe | Leu | Met | Cys | Pro | Glu | Met   | Met | Val | Gly | Met | Leu | Leu | Cys | Glu | Lys | Met | 559  |      |
| 2335 | A   | A   | T   | G   | A   | G   | T   | C     | T   | T   | A   | T   | A   | T   | C   | A   | G   | A   | T    | 2394 |
| 560  | Leu | Glu | Ala | Ile | Ser | Ile | Val | Gly   | Ala | Tyr | Val | Phe | Ser | Ser | Tyr | Thr | Gly | Tyr | 579  |      |
| 2395 | A   | C   | T   | T   | A   | T   | C   | G     | T   | T   | G   | G   | C   | --- | --- | --- | --- | --- | 2436 |      |
| 580  | Thr | Leu | Lys | Trp | Ala | Glu | Leu | Gln   | Pro | Asn | His | Ser | Arg | Gln | Asn | Thr | Asn | Glu | 599  |      |
| 2437 | G   | A   | C   | A   | T   | T   | G   | C     | A   | G   | G   | C   | G   | T   | G   | C   | A   | G   | A    | 2496 |
| 600  | Asp | Arg | Phe | Gly | Arg | Leu | Arg | Val   | Glu | Thr | Ile | Ala | Ile | Asp | Ala | Ile | Leu | Phe | 619  |      |
| 2497 | --- | --- | --- | --- | --- | --- | --- | ---   | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 2541 |      |
| 620  | Ser | Lys | Leu | Asp | Cys | Gln | Thr | Glu   | Gln | Leu | Asn | Lys | Ala | Asn | Ile | Arg | Glu | Met | 639  |      |
| 2542 | A   | A   | G   | C   | T   | T   | A   | C     | T   | C   | C   | G   | T   | C   | T   | G   | A   | A   | A    | 2601 |
| 640  | Lys | Ala | Ser | Ile | Gly | Phe | Met | Ser   | Gln | Gly | Pro | Lys | Phe | Thr | Asn | Ile | --- | --- | 658  |      |
| 2602 | A   | C   | G   | G   | A   | A   | C   | T     | G   | T   | G   | G   | T   | T   | G   | G   | G   | T   | A    | 2661 |
| 659  | Thr | Gly | Trp | Trp | Gly | Cys | Gly | Ala</ |     |     |     |     |     |     |     |     |     |     |      |      |



Db 108 AlaAlaAspGlyLeuAlaLeu-----PhePheAspAspLeuLeuSerArgAlaGlnAla 125  
QY 1762 CAACATTTATATCAGTCCATTTTACCTGACATGGTGAAATTGCACCTGTGTGCCAAAT 1821  
Db 126 ArgGlyTrpPheSerGluValValProAsnLeuAlaArgLeuLeuArgLeuProThr 145  
QY 1822 ATTTGCACCCAGCCCAATACCCTCTGAAACAGAGATGAATCATTTCT----- 1869  
Db 146 LeuLeuGluAspHisTyr-----AlaLysAlaGlyHisGlyAlaSerGlyLeu 161  
QY 1870 -----GTCACGATGTTCACAGGAACAGATCGCCAGT 1899  
Db 162 ArgValMetAlaSerGlnAspAlaGlyValValLeuLeuSerGlnGluValAlaAla 181  
QY 1900 CTTTGTAGCTAATGCTTTCTCTGACATTTCCCGCAGGAATGCCAAGATGAATCGGAG 1959  
Db 182 LeuLeuThrCysAlaLeuPheCysLeuPheProThrAlaGlyArgAlaGln----- 198  
QY 1960 TATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTT-----GAAGGA 2004  
Db 199 ---AlaCysLeuProThrIleAsnPheAspGlyLeuPheAlaProLeuIleHisAsnAla 217  
QY 2005 CGTTCATCAAGGAAACCAAGAAAACACTGAAACACTTCTGCTACTTTTCGAGAGTCACA 2064  
Db 218 ArgGlnSerGlnGlu---GlnLysValArgCysLeuValHisTyrPheGluArgValThr 236  
QY 2065 GAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGTCTT----- 2109  
Db 237 AspSerMetProThrGlyLeuValSerPheGluArgLysValLeuProArgAlaLeu 256  
QY 2110 -----GAAGATTTTCCAGAAATGGGAAAGGTGTGAAAAGCCTCTGACA 2151  
Db 257 SerAspGlyValProTyrProAspIleHisAlaTrpValAlaSerSerAlaProLeuCys 276  
QY 2152 CGCTTACACGCTCACTTACGAGGGTACCATAGAGGCAACGGCCGAGGCATGCTACAGGTG 2211  
Db 277 GlnPheThrValPheSerSerGlyPheIleGluAspGluGluGlnGluAlaLeuGlnVal 296  
QY 2212 GATTTTGCAAATCGTTTGTGAGGTTGGTGACTGGTGGCGGACTTGTACAAGAAAGA 2271  
Db 297 AspPheAlaAsnLysTyrLeuGlyGlyAlaLeuSerTrpGlyCysValGlnGluGlu 316  
QY 2272 ATCAGATTTTAAATCAATCCTGAATGATTGTTTTCACGGCTGTTCACTGAGGTGCTGGAT 2331  
Db 317 IleArgPheMetIleAsnProGluLeuIleValGlyMetLeuPheLeuSerCysMetGlu 336  
QY 2332 CACAATGAGTGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCT 2391  
Db 337 AspAsnGluAlaIleGluIlePheGlyAlaGluArgPheSerGlnTyrMetGlyTyrGly 356  
QY 2392 GAACTTATCGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAGGACGATTGGCAGCGG 2451  
Db 357 SerSerPheArgPheValGlyAspTyrLeuAspThrLysProPheAspSerMetGlyArg 376  
QY 2452 CGCTGCACGGAGATCGTTGCCATTGACGCACTTCACCTTCAGACGCTACCTCGATCAGTTT 2511  
Db 377 ArgArgThrArgIleValAlaIleAspAlaLeuAspCysProAlaArgLeu---HisTyr 395  
QY 2512 GTGCCTGAGAAAGTGACACGCTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGT----- 2565  
Db 396 GluSerGlyCysLeuLeuArgGluValAlaAsnLysAlaPheCysGlyPhePheAspGlnSer 415  
QY 2566 -----CCTGGA 2571  
Db 416 LysHisHisLeuTyrAlaLysLeuPheGlnAspLeuHisAsnLysAspPheSerSer 435  
QY 2572 GTTCTTCTGAAAATCTTTCTGAGTGCCACGGGAAACTGGGCTGTGGTGCCTTTGGG 2631  
Db 436 IleAsnSerSerGluTyrValGlyValSerThrGlyAsnTrpGlyCysGlyAlaPheGly 455  
QY 2632 GGTGACGCTAGATTAAAGCCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGAC 2691

Db 456 GlyAsnProGluIleLysSerMetIleGlnTrpIleAlaAlaSerGlnAlaLeuArgPro 475  
QY 2692 GTGGTT---TATTTACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCAC 2748  
Db 476 PheValAsnTyrTyrThrPheGluAspValSerLeuGlnArg---LeuGluGluValIle 494  
QY 2749 ACTTTCCTTACCGAGAGAAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATAC 2808  
Db 495 GlnTrpIleArgLeuHisGlyTrpThrValGlyGluLeuTrpHisMetLeuMetGluTyr 514  
QY 2809 TACAATGAA 2817  
Db 515 SerSerGln 517  
RESULT 6  
US-09-973-451-10  
; Sequence 10, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine J.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-973-451-10  
Alignment Scores:  
Pred. No.: 1.42e-26 Length: 726  
Score: 443.50 Matches: 184  
Percent Similarity: 36.62% Conservative: 124  
Best Local Similarity: 21.88% Mismatches: 278  
Query Match: 6.55% Indels: 255  
DB: 9 Gaps: 32  
US-09-302-812-5 (1-3814) x US-09-973-451-10 (1-726)  
QY 622 GAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAACTTGCAAAATACAAAGCCA 681  
Db 14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32  
QY 682 ACAGTAGGAGATGGGCAGGCCAGCAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTG 741  
Db 33 ThrMet-----LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49  
QY 742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGTCTACCAGAGAGTCTTGTGCA 801  
Db 50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSerSer 67  
QY 802 GATGTTGGTCCGAGGACATTTGGAACCTGGACCAAAA----- 837  
Db 68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87  
QY 838 -----AATGACAAACAATTG 852  
Db 88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107  
QY 853 ACTGGACAAGAAAGCAGCCTAGGTGAT-----TCGCTCCATTTGAGAAAGAAAGT 903

Db 94 LeuProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGlu 113  
QY 1480 GAGGTTCCGAAACCTTTTCCACACATTACAAAGATTGTGGGTAACAAACATGTGAAA 1539  
Db 114 ThrProProArgProTyrLysSerProGlyLys-----TipAspSerGluHisValArg 131  
QY 1540 ATGCTTGTTCGGAACAAACTTGTACCTGTGGAAGATGAGATGGTGGAGCGAACTGCA 1599  
Db 132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIle 151  
QY 1600 GGGAGTAGGTGGAGCTCATTCAGACTGCACCTTCTCAACAATTCACAGACCCCGAAGC 1659  
Db 152 AspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171  
QY 1660 TTGAAGGATGCGATTCTGAAATACAAATGAGTGGCATATTCTAAGAAATGGGACTTTACAGCT 1719  
Db 172 LeuGlnAlaAlaIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAla 191  
QY 1720 TTGGTTGATTTCTGGGATAAGGTACTTGAAGTGGCATATTCTAAGAAATGGGACTTTACAGCT 1779  
Db 192 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211  
QY 1780 ATTTTACCTGACATGGTGAATATGCACTCTGTCTGCCAAATATTTCACCCAGCCCAATA 1839  
Db 212 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerProVal 231  
QY 1840 CCACTCCTGAAACAGAAAGATGAATCATTTCTGTACAGATGTACAGAAACAGATCGCCAGT 1899  
Db 232 ProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSerCys 251  
QY 1900 CTTTGTAGCTAATGCTTTCTTCTGTCACATTTCCCGCAGCGAATGCC---AAGATGAATCG 1956  
Db 252 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSer 271  
QY 1957 GAGTATTCTAGTTACCCAGACATTAACCTCAATCGGTGTTTGAAGGACGTTTCATCAAGG 2016  
Db 272 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291  
QY 2017 AAACAGAAACAACTGAAACACTCTTCTGTCTACTTTTGAAGAGTC-----ACAGAGAAA 2070  
Db 292 ValLeuGluLysLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 311  
QY 2071 AAA-----CCTACAGGATTGGTGACATTTACAAGACAGAGT-----CTTGAA 2112  
Db 312 AspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGlu 331  
QY 2113 GATTTTCCAGATGGGAAAGGTGTGAAAGCCTCTG-----ACACGCTTACACGTCAT 2166  
Db 332 HisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351  
QY 2167 TACGAGGTACCATAGAGGCAACGCGCGGAGCATGTACAGGTGGATTTCGAAATCGT 2226  
Db 352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371  
QY 2227 TTTGTTGGAGTGGTGTGACTGCTGCGGGGACTGTACAAGAAATCAGATTTTAAATC 2286  
Db 372 TyrLeuGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheValIle 391  
QY 2287 AATCCTGAATGATTGTTTACCGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTT 2346  
Db 392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 411  
QY 2347 ATTATCAGAGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGG 2406  
Db 412 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTrp 431  
QY 2407 GCCCGAAGCCATGAAGATGGGAGTGAAGAGCAGATTGGCAGCGCGCTGCACGGAGATC 2466  
Db 432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 451  
QY 2467 GTTGCCATTGACGCACTTCACCTTCAGACGCTACCTCGATCAGTTTGTGCTGAGAAAGTG 2526  
Db 452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisGlnTyrArgGluAspLeuMet 471

QY 2527 AGACGTGAGCTTAACAAGGCTTACTGCGGATTCTCCCGT-----CCT 2568  
Db 472 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 491  
QY 2569 GGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTT 2628  
Db 492 Gly-----ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 503  
QY 2629 GGGGTGACGCTAGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGGCTGAACGT 2688  
Db 504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523  
QY 2689 GACGTGTTTATTTTACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCAC 2748  
Db 524 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMetTrp 543  
QY 2749 ACTTTCCTTACCGAGAGGAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATAC 2808  
Db 544 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSer 562  
QY 2809 TACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA-----GAC 2850  
Db 563 TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582  
QY 2851 ATCAAGCTTTATCCATTTCATATACCATGCTGTTGAGTCAAGTGCAGAGACCACACTGACATG 2910  
Db 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599  
QY 2911 CCAGGACAGAGGCA 2925  
Db 600 ProGlyGluGlyAla 604

RESULT 5

US-10-425-114-60000  
; Sequence 60000, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 60000  
; LENGTH: 546  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3689-262-E12\_FLI.pep  
US-10-425-114-60000

Alignment Scores:  
Pred. No.: 1.44e-37 Length: 546  
Score: 579.00 Matches: 150  
Percent Similarity: 49.44% Conservative: 69  
Best Local Similarity: 33.86% Mismatches: 154  
Query Match: 8.55% Indels: 70  
DB: 12 Gaps: 12

US-09-302-812-5 (1-3814) x US-10-425-114-60000 (1-546)

QY 1660 TTGAAGGATGCGATTCTGAAATACAAATGTGGCATATTCT-----AAG 1701  
Db 88 LeuAlaAspAlaLeuThrAspLeuArgLeuAlaLeuProAlaLeuProArg 107  
QY 1702 AATGGGACTTTACAGCTTTGTTGATTCTGGGATAAGTACTTGAAGAAGCAGAGGCC 1761

Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
QY 1561 TTGTACCCCTGTGGAAGATGAGAAATGGTGAAGCACTGCAGGGAGTAGGTGGAGCTCATT 1620  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
QY 1621 CAGACTGCACCTTCTCAACAAATTCACACGACCCCGAAGCTTGAAGGATGCGATTCTGAAA 1680  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
QY 1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740  
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
QY 1741 GTACTTTGAAGAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGCACATGGTGAAA 1800  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
QY 1801 ATTGCACCTCTGTGCGCAAAATATTGTCACCCAGCCCAATACCACTCTCTGAAACAGAGATG 1860  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
QY 1861 AATCATTTCTGCAGATGTCACAGGAACAGATGCCAGTCTTTTAGCTAATGCTTCTTCTTC 1920  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
QY 1921 TGCACATTTCCCGACGGAATGCCAAGATGAATCGGAGTATTCTAGTTACCCAGACATT 1980  
Db 641 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTrpProAspIle 660  
QY 1981 AACTTCAATCGTGTGTTGAAGGACGTTTCATCAAGGAAACAGAAAACCTGAAAACACTC 2040  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
QY 2041 TTCTGTCTACTTTGGAAGAGTCACAGAGAAAACCTACAGGATTGGTGACATTTACAAGA 2100  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAGGTGTGAAAAGCCTCTGACAGCTTACAC 2160  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
QY 2161 GTCACCTTACGAGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGTGGATTGTCAC 2220  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2221 AATCGTTTTGTGGAGGTGGTGTGACTGGTGGGGAAGCTTGTACAAGAGAAATCAGATT 2280  
Db 741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
QY 2281 TTAATCAATCTGAATTTGATTGTTTCCAGGCTGTTCACTGAGGTGCTGGATCACAATGAG 2340  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAACTTAT 2400  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2401 CGTTGGGCGCGAAGCCATGAAGATGGGAGTGAAAGGACGATTGGCAGCGCGCTGCACG 2460  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820  
QY 2461 GAGATCGTTGCCATTGACGCACCTTCACTCAGACGCTACCTCGATCAGTTTGTGCTGAG 2520  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2521 AAAGTGACAGCTGAGCTTAACAAGGCTTACTGCGGATTCTCTCGCTCGGAGTTCTTCT 2580  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2581 GAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGCTGTGGTGCCTTTGGGGTGACGCT 2640  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

QY 2641 AGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGAGTGGTTTAT 2700  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAAGTTATTGCTTAGATACTACAATGAAGAA 2820  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuArgTyrTyrAsnGluGlu 940  
QY 2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTATATACCATGCT 2880  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 2881 GTTGAGTCAAGTGCAGAGACCACCTGACATGCCAGGACAGAGGAGGCAGGC 2928  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976  
RESULT 4  
US-09-973-451-8  
; Sequence 8, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-973-451-8  
Alignment Scores:  
Pred. No.: 1.65e-77 Length: 768  
Score: 1074.50 Matches: 244  
Percent Similarity: 58.12% Conservative: 96  
Best Local Similarity: 41.71% Mismatches: 194  
Query Match: 15.86% Indels: 51  
DB: 9 Gaps: 15  
US-09-302-812-5 (1-3814) x US-09-973-451-8 (1-768)  
QY 1249 AGAATTTCCAAGTCAGAGGACAGAGAAAGAAACAATGTGAAGTCAGACATCAAGAACA 1308  
Db 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGlu----- 62  
QY 1309 GAAAGGAAGATTCCAAATACATCCACCTAACCTCCCTCCA-----GAGAAGAGTGG 1362  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74  
QY 1363 CTGGGAACCTCTATTGAGGAATGAGAAAAATGCCTCGTGTGGGATCCATTGCTCTCC 1422  
Db 75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93  
QY 1423 TTAAGACCA---TCTGCAAGTCACACAGTACTGTTCGGGTAGACCTTCTGAGAGCAGGA 1479



Db 940 CysArgAsnCysSerThrProGlyProAspIleLeuTyrProPheIleTyrHisAla 959

QY 2881 GTTGAGTCAAGTGCAGACCACTGACATGCCAGACAGAAAGCAGGCACC 2931

Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 3

US-09-973-451-2

; Sequence 2, Application US/09973451

; Patent No. US20020132328A1

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE

; TITLE OF INVENTION: (PARG) ENZYMES,

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/973,451

; PRIOR FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: US/09/302,812

; PRIOR FILING DATE: 1999-04-30

; PRIOR APPLICATION NUMBER: 60/083,768

; PRIOR FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 2

; LENGTH: 977

; TYPE: PRT

; ORGANISM: Bos taurus

; FEATURE:

US-09-973-451-2

Alignment Scores:

|                        |         |               |     |
|------------------------|---------|---------------|-----|
| Pred. No.:             | 0       | Length:       | 977 |
| Score:                 | 4332.50 | Matches:      | 820 |
| Percent Similarity:    | 89.86%  | Conservative: | 57  |
| Best Local Similarity: | 84.02%  | Mismatches:   | 90  |
| Query Match:           | 63.97%  | Indels:       | 9   |
| DB:                    | 9       | Gaps:         | 5   |

US-09-302-812-5 (1-3814) x US-09-973-451-2 (1-977)

QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCACGAAA---GCGCGCTGGGGCGCGCTGGA 84

Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20

QY 85 ACTTCTGCGCGACTGCCTCGGACTCCCGGAGCTCCCTGGCAGGCGAGCGGTGTTCTC 144

Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40

QY 145 GACCCCAAGGAGCTCCCGTCCAGTTCAGGTCCCTCCGTCCTCGCCAGCGCTGCTCG 204

Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

QY 205 GGGCGGGCGGACCGCACAGAGGCAACGCCACCTCGTTGTTTTCACAAACAAAGACTATT 264

Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

QY 265 ACTACTTGATGGATACTAAAGGACCCAGACAGCTGAATCAGAA-----AGTAAA 315

Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

QY 316 GAAACAACATACAAAGAAATTGACTCCATGATGAGTCTGTGTCAGAAAGACTATT 375

Db 101 GluAsnAsnThrArgGluGluSerMetSerSerValGlnLysAspAsnPheTyr 120

QY 376 CCACATAAGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCACCA 435

Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

QY 436 GAAAAGAGTTACAGTATTTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGCAGAAT 495

Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

QY 496 GAAGGGAAGCATGCAGAAACAGCTTTTGGAGAGTGCCTCCCGGGGACTCCGCTACCA 555

Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

QY 556 AAGCAGCTTAGTAATGCTAAACATTGGTCAGTCACCCACACTGATGACACAGTGACACA 615

Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

QY 616 GATCATGAAGAAGACAGACAAATCAGCAGTTTCTTACACCTATAAAACTTGCATAATACA 675

Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

QY 676 AAGCCAACAGTAGGAGATGGGCAG-----GCCAGAAGCAACTGTAAGTGCAGTGA 726

Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

QY 727 TCTCGCCAGCTCTGTGAAAGACTGTACAGCTGTCAACAGGAGGAGGTGGATGTCTACCA 786

Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260

QY 787 GAGAGTCCCTTTGTGATGTTGGTCCGAGGACATTTGGAACCTGGACCAAAATGACAAC 846

Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

QY 847 AAATTGACTGACAAAGAAAGCAGCCTAGTGTATTCGCCTCCATTTGAGAAAAGAGTGAG 906

Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

QY 907 CCTGAGTCACCAATGGATGTAGACAACCTCGAGAAAACAGTTGTCAAGATTCCAGAACGAGAT 966

Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

QY 967 GAAGAAACAAGTCCAGTCTTTTGATGAGCAAGATGATCGTTCCTCC---CAAACAGCAAAAT 1023

Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

QY 1024 AAATTTCAGTTGCCAAGCAAGAGAGTGTGATGGCGATCTTAGGAAAACGTTATTGACT 1083

Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

QY 1084 AAGGGAAGTGAAGTTAGATTGCAATTCCTCAATTTGAA---GGAGAAAATAATGCTGGGACC 1140

Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

QY 1141 AGTGACTTAATGCCAAGCCATCTGGAACTCTTCTAGCCTTAATGATGAGTGTAGAAGT 1200

Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

QY 1201 TCCAAGCAGCATGGAAAAGGATTCTTAAATTTACAGATCATTTTCATGAGAAATTTCCAAG 1260

Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

QY 1261 TCAGAGGACAGAAAGAAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATT 1320

Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

QY 1321 CCAAAATACATCCACCTAACCTCCCTCCAGAGAAGAGTGGTGGGAACTCCTATTGAG 1380

Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

QY 1381 GAAATGAGAAAATGCCCTCGGTGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGT 1440

Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

QY 1441 CACACAGTGACTGTTCCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCGAAACCTTTTCCA 1500

Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

QY 1501 ACACATTACAAAGATTGTGGGATAACAAACATGTGAAAATGCTTGTTCGGAAACAAAAC 1560



QY 676 AAGCCAAACAGTAGGAGATGGCGAGCGCAGAG-----AGCAACTGTAAGTGCAGTGA 726  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCA 786  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
QY 787 GAGAGTCTTTGTTCAGATGTTGGTCCGAGGACATTGGAACCTGGACCAAAAAATGACAC 846  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 847 AAATTGACTGGACAAGAACGACCTAGGTGATTGGCTCCATTGAGAAAGAAAGTGAG 906  
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299  
QY 907 CCTGAGTACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAAGCAGAT 966  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAAAT 1023  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
QY 1024 AAACCTTCAAGTTGCCAAGCAAGAGAAAGCTGATGGCGATCTTAGGAAACGGTATTGACT 1083  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1084 AAGGGAAGTGAAGTTAGATTGCAATTTCAATTTGAA---GGAGAAATAATGCTGGGACC 1140  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAAATGCCAAGCCATCTGTGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAAGCAGCATGGAAAAGGATTTCTAAAATTACAGATCATTTTCATGAGAAATTTCCAAG 1260  
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAAGAAAGAAACAATGTGAAGTCAGACATCAAAAGAAACAGAAAGGAAGATT 1320  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1321 CCAAAATACATCCCACTTAACCTCCCTCCAGAGAAAGAGTGGCTGGGAACCTCCTATTGAG 1380  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1381 GAAATGAGAAAAATGCTCGGTGTGGATCCATTTCCTTTCCTTAAGACCCTCTGCAAGT 1440  
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACACAGTGACTGTTCCGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1501 ACACATTACAAAGATTGTGGGATAACAAACATGTGAAAATGCCCTTGTTCGGAAACAAAAC 1560  
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1561 TTGTACCTGTGGAAATGAGAAATGGTGGCGAACTGCGAGGAGTAGGTGGAGCTCATT 1620  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1621 CAGACTGCACCTTCTCAACAAATTCACACGACCCCAAGAACTTGAAGGATGCGATTCTGAAA 1680  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTCTGGGATAAG 1740  
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1741 GTACTTGAAGAGCAGAGCGGCCCAACATTTTATATCAGTCCATTTTACCTGACATGGTGAAA 1800

Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCACCTCTGTCTGCCAAATATTTGCACCCAGCCAAATACCACCTCTCTGAAACAGAGATG 1860  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATTTCTGTACGATGTACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuAlaAsnAlaPhePhe 639  
QY 1921 TGCACATTTCCCCGACCGAATGCCAAGATGAAATCGAGTATTCTAGTTACCCAGACATT 1980  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 1981 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACAGAAAAAACTGAAAAACACTC 2040  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2041 TTCTGCTACTTTTCGAAGAGTACAGAGAAAAAACTACAGGATTGGTGACATTTACAAGA 2100  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2101 CAGAGTCTTGAAGATTTCCAGAAATGGGAAAGGTGTGAAAAAGCCTCTGACACGCTTACAC 2160  
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2161 GTCACCTTACGAGGTTACCATAGAGGCAACCGCCGAGGCATGCTACAGGTGGATTTTGCA 2220  
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2221 AATCGTTTGTGAGGTGTGTGACTGGTGGCGGACTTGTACAAGAAGAAATCAGATTT 2280  
Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
QY 2281 TTAATCAATCCTGAATTGATTGTTTTCACGGCTGTTTCACTGAGGTGCTGGATCAATGAG 2340  
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAsnGlu 779  
QY 2341 TGTCTTATTATCAGAGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2401 CGTTGGCCCGAAGCCATGAAGATGGAGTGGAGTGAAGGACGATTGGCAGCGCGCTGCACG 2460  
Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2461 GAGATCGTTGCCATTGACGCACCTTCACCTTCAGACGCTACCTCGATCAGTTTGTCCCTGAG 2520  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2521 AAAGTGAGACGTGAGCTTAAACAAGGCTTACTGCGGATTCTCCGCTCCGAGTTCTTCT 2580  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2581 GAAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2641 AGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGTTTAT 2700  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCTCTTACC 2760  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACATCAATGAAGAA 2820  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 2821 TGCAGAAAACCTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 2880

QY 2008 TCATCAAGGAACCCAGAAAACACTGAAACACTCTTCTGCTACTTTTCGAAGAGTCACAGAG 2067  
Db 661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu 680  
QY 2068 AAAAAACCTACAGGATTGGTGACATTTCACAGACAGAGTCTTGAAGATTTTCCAGATGG 2127  
Db 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp 700  
QY 2128 GAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACCTTACGAGGGTACCATAGAAGGC 2187  
Db 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGGCCGAGGCATGCTACAGGTGGATTGTCAAATCGTTTGTGGAGGTGGTGTGACT 2247  
Db 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGACTTGTACAGAAGAAATCAGATTTTAAATCAATCCTGAATTGATTGTTCA 2307  
Db 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACAG 2367  
Db 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCGAAGCCATGAAGATGG 2427  
Db 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCAC 2487  
Db 801 SerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACCTCGATCAGTTTGTGCTGAGAAAGTGAAGACGTGAGCTTAACAAGGCT 2547  
Db 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840  
QY 2548 TACTGCGGATTCTCCGCTCGGAGTTCCTTCTGAAATCTTCTGCAAGTGGCCACGGGA 2607  
Db 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGCTGTGGTCTTGGGGTGACGTGTTTATTTACCTTTGGGACTCAGAGTTGATG 2667  
Db 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
QY 2668 GCAGCTGCTGCGGCTGAACGTGACGTGGTGTATTTACCTTTGGGACTCAGAGTTGATG 2727  
Db 881 AlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGAGACATTTACACATGCACACTTTTCCCTTACCGAGAGGAAGCTGGATGTTGGAAGTG 2787  
Db 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
QY 2788 TACAAGTTATTGCTTAGATACTACAATGAAGAAATGCAGAAACTGTTCCACCCCTGGACCA 2847  
Db 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTATCCATTATATACCATGCTGTTGAGTCAAGTGCAGAGACCCTGAC 2907  
Db 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
QY 2908 ATGCCAGGACAGAGCGCAGGCACC 2931  
Db 961 MetProGlyGlnLysAlaGlyThr 968

RESULT 2

US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4  
Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 64.89% Indels: 10  
DB: 9 Gaps: 6

US-09-302-812-5 (1-3814) x US-09-973-451-4 (1-976)

QY 28 ATGAGTGGGGCCCGGCTGGAGCCCTGCACGAAAGCG---CGCTGGGGCGCGCTGGA 84  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19  
QY 85 ACTTCTGCGCGGACTGCTCGGACTCCCGGAGCTCCCTGGCAGGAGCGGTGTTCTC 144  
Db 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39  
QY 145 GACCCCAAGGACGCTCCGTCAGTTCCAGGTCCCTCCGTCCTCGCCAGCTGCTCTCG 204  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGGCGGACCGCACAGAGGCAACGCCACCTCGTTTGTGTTTCAACAAAGACTATT 264  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 265 ACTACTTGATGATGATACTAAAGGACCCCAAGACAGCTGAATCAGAA-----AGTAA 315  
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 316 GAAACAACAATAACAAGATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTAC 375  
Db 100 GluAsnAsnAsnThrArgIleGluSerMetSerValGlnLysAspAsnPheTyr 119  
QY 376 CCACATAAGGTGGAATAATTTGAAAATGTTCTCTCAGCTAAATCTTGATAAATCACCACA 435  
Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 436 GAAAGAGTTCACAGTATTGTAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGAGCTCCGCTACCA 555  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 556 AAGCAGCTTAGTAATGCTAATGTTGTCAGTCCAGTCACCCACACTGATGACCACAGTGACACA 615  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 616 GATCATGAAGAAGACAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACA 675  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219

```

; ORGANISM: Mus musculus
; FEATURE:
US-09-973-451-6

Alignment Scores:
Pred. No.: 0
Score: 5164.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 76.24%
DB: 9

US-09-302-812-5 (1-3814) x US-09-973-451-6 (1-968)

QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCACGAAAGCGCGCTGGGGCGCGCTGGAAC 87
Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20

QY 88 TCTGCGCGACTGCCCTCGACTCCCGAGCTTCCCTGGCAGGCAGAGCGGTCTCGAC 147
Db 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeuAsp 40

QY 148 CCCAAGGACGCTCCCGTCCAGTTCAGGTCCCTCCGTCCTGCCAGCCTGCGTCTCGGG 207
Db 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60

QY 208 CGGGCGGACCGCACAGAGGCAACGCCACCTCGTTTGTGTTTCAACAAAGACTATTACT 267
Db 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80

QY 268 ACTTGGATGGATACATAAGGACCCCAAGACAGCTGATCAGAAAGTAAAGAAACAACAT 327
Db 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100

QY 328 ACAAGAATTGAAATGTCCTCAGCTAAATCTTGATGAAATACCCACAGAAAGAGTTCA 447
Db 101 ThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyrProHisLysVal 120

QY 388 GAAAAATTGAAATGTCCTCAGCTAAATCTTGATGAAATACCCACAGAAAGAGTTCA 447
Db 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer 140

QY 448 CAGTATTGAACCAACAGCAGACTGCGAGTGTGCAAGTGGCAGAAATGAAGGAAGCAT 507
Db 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160

QY 508 GCAGAACAGCTTTTGGCAAGTGAGCTCCCGCGGGACTCCGCTACCAAGCAGCTTAGT 567
Db 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180

QY 568 AATGCTAACATTCAGTCAGTCACCCACACATGATGACACAGTGACACAGATCATGAAGA 627
Db 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200

QY 628 GACAGAGACAATCAGCAGTTTCTTACACCTATATAAACTTGCAATACAAAGCCCAACAGTA 687
Db 201 AspArgAspAsnGlnGlnPheLeuThrProLysLeuAlaAsnThrLysProThrVal 220

QY 688 GGAGATGGGACGCCAGAGCAACTGTAGTCAGTGGATCTCGCCAGTCTGTGAAAGAC 747
Db 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240

QY 748 TGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACAGAGAGTCCTTGTTCAGATGTT 807
Db 241 CysThrGlyCysGlnGlnGluGluValAspValLeuProGluSerProLeuSerAspVal 260

QY 808 GGTGCCGAGGACATTGGAACCTGGACCAAAAAATGACAAACAATGACTGGACAAGAAAGC 867
Db 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280

QY 868 AGCCTAGGTGATTCGCTCCATTGTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTA 927
Db 281 SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal 300
```

```

QY 928 GACAACTCGAGAAACAGTTGTCAAGATTTCAGAAAGCAGATGAAGAAACAAAGTCCAGTCTTT 987
Db 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320

QY 988 GATGAGCAAGATGATCGTTCCTCCCAACACAGCAATAAATACTTTCAAGTTGCCAAGCAAGA 1047
Db 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340

QY 1048 GAAAGCTGATGGCGATCTTAGGAAACCGTATTGTAAGGGAAGTGAAGTTAGATTGCAT 1107
Db 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360

QY 1108 TTCCAATTTGAAGGAGAAAAATAATGCTGGGACCAGTACCTTAAATGCCAAGCCATCTGGA 1167
Db 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380

QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAGTGTAGAGTTCAGAGCAGCATGGAAAAAGGGATTCT 1227
Db 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400

QY 1228 AAAATTACAGATCATTTTCATGAGAAATTTCCAGTCCAGAGCAGACAGAAAGAAACAATGT 1287
Db 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420

QY 1288 GAAAGTCAGACATCAAAAGAACAGAAAGAGAGATTCCAAAATACATCCCACTAACCTCCCT 1347
Db 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440

QY 1348 CCAGAGAAGAGTGGCTGGGAACCTCTTATGAGGAAATGAGAAAAATGCCTCGGTGTGGG 1407
Db 441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460

QY 1408 ATCCATTGTCCTTCTTAAGACCATCTGCAAGTCACACAGTGCATGTTTCGGGTAGACCTT 1467
Db 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480

QY 1468 CTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAGATTTGTGGGATAAC 1527
Db 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500

QY 1528 AAACATGTGAAAATGCCCTTGTTCGGAAACAAACTTGTACCCCTGTGGAAGATGAGAATGGT 1587
Db 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520

QY 1588 GAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCATCTCTCAACAAAATTCACA 1647
Db 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540

QY 1648 CGACCCCGAGAACTTGAAGGATGCGATTCTGAAATACAAATGTGGCATATTTAAGAAATGG 1707
Db 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560

QY 1708 GACTTTACAGCTTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAGCAGAGGCCCAACAT 1767
Db 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGlnHis 580

QY 1768 TTATATCAGTCCATTTTACCTGACATGGTGAATAATTCACCTCTGTCTGCCAAATATTTGC 1827
Db 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600

QY 1828 ACCAGCCAATACCACTCCTGAAACAGAAAGATGAATCATCTTGTTCAGATGTACAGGAA 1887
Db 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620

QY 1888 CAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCGACGGAATGCCAAG 1947
Db 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640

QY 1948 ATGAAATCGGAGTATTCTAGTTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGT 2007
Db 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: May 26, 2004, 17:10:29 ; Search time 158.584 Seconds  
(without alignments)  
13416.302 Million cell updates/sec

Title: US-09-302-812-5  
Perfect score: 6773  
Sequence: 1 gggggactgtgtgctgctgggg.....aatcattgtcagaaaaaaa 3814

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 2298626

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/6333148/runat\_26052004\_150056\_5800/app\_query.fasta\_1.12437  
-DB=Published Applications\_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=6333148 @CGN 1 1 79 @runat\_26052004\_150056\_5800  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| -----      |       |             |        |    |             |

|    |        |      |      |    |                      |                   |
|----|--------|------|------|----|----------------------|-------------------|
| 1  | 5164   | 76.2 | 968  | 9  | US-09-973-451-6      | Sequence 6, Appli |
| 2  | 4395   | 64.9 | 976  | 9  | US-09-973-451-4      | Sequence 4, Appli |
| 3  | 4332.5 | 64.0 | 977  | 9  | US-09-973-451-2      | Sequence 2, Appli |
| 4  | 1074.5 | 15.9 | 768  | 9  | US-09-973-451-8      | Sequence 8, Appli |
| 5  | 579    | 8.5  | 546  | 12 | US-10-425-114-60000  | Sequence 60000, A |
| 6  | 443.5  | 6.5  | 726  | 9  | US-09-973-451-10     | Sequence 10, Appl |
| 7  | 220.5  | 3.3  | 200  | 12 | US-10-424-599-184988 | Sequence 184988,  |
| 8  | 176    | 2.6  | 6642 | 15 | US-10-369-493-5013   | Sequence 5013, Ap |
| 9  | 175    | 2.6  | 180  | 12 | US-10-424-599-156445 | Sequence 156445,  |
| 10 | 172.5  | 2.5  | 300  | 12 | US-10-424-599-233915 | Sequence 233915,  |
| 11 | 172    | 2.5  | 1489 | 15 | US-10-115-482-18     | Sequence 18, Appl |
| 12 | 169.5  | 2.5  | 1297 | 12 | US-10-282-122A-71692 | Sequence 71692, A |
| 13 | 167    | 2.5  | 31   | 9  | US-09-973-451-11     | Sequence 11, Appl |
| 14 | 167    | 2.5  | 1169 | 12 | US-10-282-122A-71179 | Sequence 71179, A |
| 15 | 166    | 2.5  | 1210 | 14 | US-10-205-823-264    | Sequence 264, App |
| 16 | 163    | 2.4  | 691  | 15 | US-10-108-260A-4345  | Sequence 4345, Ap |
| 17 | 163    | 2.4  | 2677 | 15 | US-10-144-194A-22    | Sequence 22, Appl |
| 18 | 162.5  | 2.4  | 4723 | 15 | US-10-359-012-8      | Sequence 8, Appli |
| 19 | 161    | 2.4  | 748  | 9  | US-09-864-761-43244  | Sequence 43244, A |
| 20 | 160.5  | 2.4  | 1462 | 14 | US-10-287-218-17     | Sequence 17, Appl |
| 21 | 159.5  | 2.4  | 1501 | 9  | US-09-924-154-17     | Sequence 17, Appl |
| 22 | 159    | 2.3  | 2697 | 15 | US-10-144-198-12     | Sequence 12, Appl |
| 23 | 155.5  | 2.3  | 917  | 15 | US-10-221-658-4      | Sequence 4, Appli |
| 24 | 152.5  | 2.3  | 1404 | 9  | US-09-811-045A-1     | Sequence 1, Appli |
| 25 | 151    | 2.2  | 913  | 13 | US-10-117-604-4      | Sequence 4, Appli |
| 26 | 150    | 2.2  | 447  | 10 | US-09-794-422-44     | Sequence 44, Appl |
| 27 | 150    | 2.2  | 1115 | 13 | US-10-153-273-2      | Sequence 2, Appli |
| 28 | 150    | 2.2  | 1881 | 14 | US-10-032-585-7646   | Sequence 7646, Ap |
| 29 | 149.5  | 2.2  | 1973 | 16 | US-10-197-824-34     | Sequence 34, Appl |
| 30 | 149.5  | 2.2  | 2047 | 12 | US-10-346-863-2      | Sequence 2, Appli |
| 31 | 149.5  | 2.2  | 3863 | 16 | US-10-197-824-7      | Sequence 7, Appli |
| 32 | 148    | 2.2  | 1358 | 15 | US-10-369-493-1658   | Sequence 1658, Ap |
| 33 | 148    | 2.2  | 1389 | 12 | US-10-282-122A-70151 | Sequence 70151, A |
| 34 | 148    | 2.2  | 2063 | 12 | US-09-918-715-204    | Sequence 204, App |
| 35 | 147    | 2.2  | 823  | 12 | US-10-232-759A-10    | Sequence 10, Appl |
| 36 | 146.5  | 2.2  | 453  | 15 | US-10-440-464-115    | Sequence 115, App |
| 37 | 145.5  | 2.1  | 498  | 12 | US-10-424-599-196154 | Sequence 196154,  |
| 38 | 145.5  | 2.1  | 2442 | 13 | US-10-109-886-10     | Sequence 10, Appl |
| 39 | 145    | 2.1  | 1139 | 10 | US-09-820-843A-15    | Sequence 15, Appl |
| 40 | 145    | 2.1  | 1139 | 12 | US-10-282-122A-63564 | Sequence 63564, A |
| 41 | 144.5  | 2.1  | 747  | 12 | US-10-425-114-67803  | Sequence 67803, A |
| 42 | 144    | 2.1  | 26   | 9  | US-09-973-451-21     | Sequence 21, Appl |
| 43 | 144    | 2.1  | 29   | 9  | US-09-973-451-12     | Sequence 12, Appl |
| 44 | 144    | 2.1  | 1189 | 15 | US-10-369-493-22159  | Sequence 22159, A |
| 45 | 143.5  | 2.1  | 1163 | 12 | US-10-221-278-197    | Sequence 197, App |

ALIGNMENTS

RESULT 1  
US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT



|    |      |  |  |  |  |  |      |
|----|------|--|--|--|--|--|------|
| Db | 432  | -----AlaArgProAspSerProPheSer-----                           |  |  |  |  | 440  |
| QY | 1987 | AATCGGTTGTTGAAGGACGTTTCATCAAGGAAACAGAAAACTGAAAACTCTTCTGC     |  |  |  |  | 2046 |
| Db | 441  | CysArgIleLeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThr |  |  |  |  | 460  |
| QY | 2047 | TACTTTCGAAGAGTCACAGAGAAAAACCTACAGGATTGGTGACATTT-----ACAAGA   |  |  |  |  | 2100 |
| Db | 461  | TyrPheAspLysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLys |  |  |  |  | 480  |
| QY | 2101 | CAGAGTCTTGAAGATTTT---CCAGAATGG---GAAAGGTGTGAAAAGCCTCTGACACGC |  |  |  |  | 2154 |
| Db | 481  | MetAspLysAspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGlu |  |  |  |  | 500  |
| QY | 2155 | TTACACGTCACTTACGAGGTTACCATAGAGGCAACGGCCGAGGCATGCTACAGGTGGAT  |  |  |  |  | 2214 |
| Db | 501  | ValGluPhePheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAsp |  |  |  |  | 519  |
| QY | 2215 | TTTGCAAATCGTTTGTGGAGGTGGTGTGACTGGTGGCGGACTTGTAAGAAGAAATC     |  |  |  |  | 2274 |
| Db | 520  | PheAlaAsnGluHisLeuGlyGlyValLeuAsnHisGlySerValGlnGluGluIle    |  |  |  |  | 539  |
| QY | 2275 | AGATTTTAAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCAC |  |  |  |  | 2334 |
| Db | 540  | ArgPheLeuMetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGln |  |  |  |  | 559  |
| QY | 2335 | AATGAGTGTCTTATTATACAGAGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAA |  |  |  |  | 2394 |
| Db | 560  | LeuGluAlaIleSerIleValGlyAlaIleValPheSerSerTyrThrGlyTyrGlyHis |  |  |  |  | 579  |
| QY | 2395 | ACTTATCGTTGGGCC-----CGAAGCCATGAAGATGGGAGTGAA---AAG           |  |  |  |  | 2436 |
| Db | 580  | ThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArg |  |  |  |  | 599  |
| QY | 2437 | GACGATTGGCAGCGGGCTGCACGGAGATCGTTGCCATTGACGCACCTTCACTTCAGACGC |  |  |  |  | 2496 |
| Db | 600  | AspArgPheGlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGly |  |  |  |  | 619  |
| QY | 2497 | -----TACCTCGATCAGTTTGTGCGCTGAGAAAGTGAGCGTGAGCTTAAC           |  |  |  |  | 2541 |
| Db | 620  | SerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLys |  |  |  |  | 639  |
| QY | 2542 | AAGGCTTACTGCGGATTCTCCGTCCTGGAGTTCCTTCTGAAAAATCTTCTGCAGTGGCC  |  |  |  |  | 2601 |
| Db | 640  | LysAlaSerIleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleVal |  |  |  |  | 658  |
| QY | 2602 | ACGGGAAACTGGGGCTGTGGTGCCCTTGGGGCTGACGCTAGATTAAAAGCCTTAATACAG |  |  |  |  | 2661 |
| Db | 659  | ThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGln |  |  |  |  | 678  |
| QY | 2662 | ATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTCACCTTTGGGGACTCAGAG |  |  |  |  | 2721 |
| Db | 679  | ValIleAlaAlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGlu |  |  |  |  | 698  |
| QY | 2722 | TTGATGAGAGACATTTACAGCATGCACACTTTCCTTACCAGAGAGGAGCTGGATGTTGGA |  |  |  |  | 2781 |
| Db | 699  | LeuAlaAlaLysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGly |  |  |  |  | 718  |
| QY | 2782 | AAA  |  |  |  |  | 2784 |
| Db | 719  | Lys  |  |  |  |  | 719  |

Db 719 Lys 719  
RESULT 15  
US-09-511-507-10  
; Sequence 10, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-507-10  
Alignment Scores:  
Pred. No.: 1.27e-33 Length: 726  
Score: 443.50 Matches: 184  
Percent Similarity: 36.62% Conservative: 124  
Best Local Similarity: 21.88% Mismatches: 278  
Query Match: 6.55% Indels: 255  
DB: 4 Gaps: 32  
US-09-302-812-5 (1-3814) x US-09-511-507-10 (1-726)  
QY 622 GAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATATAAACTTGCAAAATACAAAGCCA 681  
Db 14 GlnAspGluLysAspTyrGluAspTyrVal--GlyValGlyPheAlaHisGlnValPro 32  
QY 682 ACAGTAGGAGATGGCAGGCCAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTG 741  
Db 33 ThrMet-----LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49  
QY 742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAAGAGAGTCTTGTCA 801  
Db 50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSerSer 67  
QY 802 GATGTTGGTCCGAGGACATTGGAACTGGACCAAAA----- 837  
Db 68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87  
QY 838 -----AATGACAAACAAATTG 852  
Db 88 CysGluAsnLeuThrGluThrLysLysIleSerAsnIleGluSerLeuAspAsnValThr 107  
QY 853 ACTGGACAAGAAAGCAGCCTAGGTGAT-----TCGCCTCCATTGTGAGAAAGAAAGT 903  
Db 108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127  
QY 904 GAGCCTGAGTCACCAATGGATGTA----- 927  
Db 128 AsnAsnLysSerAsnIleAspValAlaIleAsnSerAspGluAspAspGluLeuValLeu 147  
QY 928 ---GACAACTCGAGAAACAGTTGTCAAGATTCAGAAAGCAGATGAAGAAACAAAGTCCAGTC 984  
Db 148 GluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnGlnLeuSerGlnAspLeu 167  
QY 985 TTT---GATGAGCAAGAT----- 999  
Db 168 PheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrThrGln 187

QY 1000 -----GATCGTTCTCCAAAACAGCAAAATAAACTTTCAAGTTGCCAAGCAAGA 1047  
Db 188 LeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThr 207  
QY 1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107  
Db 208 GluAlaAspSer----- 211  
QY 1108 TTCCAAATTTGAAGGAGAAATAATGCTGGGACCAGTACCTTAAATGCCAAGCCATCTGGA 1167  
Db 212 ---ThrPheValGlyGluAspSerLysAlaThrLysThrValArgThrSerSerSer 230  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAGCAGCATGGAAAAGG----- 1221  
Db 231 PheLeuSerThrValSerThr---CysGluAlaProAlaLysGlyArgAlaArgMetTyr 249  
QY 1222 GATTCTAAAATTACAGATCATTTTCATGAGAAATTTCCAAAGTCAGAG----- 1266  
Db 250 GlnLysGluLeuGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnPro 269  
QY 1267 -----GACAGAAGAAAGAAACAATGTGAAGTCAGACATCAAAGA 1305  
Db 270 AspLeuAsnLysValAspProAspArgAsnTyrArgTyrCysThrIle----- 285  
QY 1306 ACAGAAAGGAAGATTCCAAAATACATCCACCTAACCTCCCTCCAGAGAAAGTGGCTG 1365  
Db 286 -----ProAsnPheProAlaSerGln----- 292  
QY 1366 GGAACCTCTATTGAGGAAATGAGAAAAATGCTCGGTGGGATCCATTTGCCCTTCCTTA 1425  
Db 293 GlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeu----- 307  
QY 1426 AGACCATCTGCAAGTCACACAGTCACTGTTCCGGGTAGACCTTCTGAGAGCAGGAGGTT 1485  
Db 307 ----- 307  
QY 1486 CCGAAACCTTTTCCAAACACATTAACAAAGATTGTGGGATAACAAACATGTGAAATGCCT 1545  
Db 308 -----ProGlnArgTrpArgGluPhe----- 314  
QY 1546 TGTTCCGAACAAAACCTTGTAACCTGTGGAAGATGAGAATGGTGAGCGA----- 1593  
Db 315 -----AspSerArgGlyArgArgArgAspSerTyr 324  
QY 1594 -----ACTGCAGGG----- 1602  
Db 325 PheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPhe 344  
QY 1603 -----AGTAGTGGGAGCTCATTCAGACTGCACCTTCTCAACAAA 1641  
Db 345 MetPheValGlyLeuLeuHisAsnMetTrpGluPhe----- 356  
QY 1642 TTCACAGACCCAGAACCTTGAAGGATGCGATTCTGAATACAATGTGGCA----- 1692  
Db 357 -----AspProAspIleThrTyrLysLeuProAlaLeuGlu 368  
QY 1693 -----TATTCTAAGAAATGGGACTTTTACAGCTTTGGTGTGATTTCTGGGATAAGGTACTT 1746  
Db 369 MetTyrTyrLysGluMetSerGluLeuValGlyArgGluGluValLeuGluLysPheAla 388  
QY 1747 GAAGAAGCAGAGGCCCCAACATTTATATCAGTCCATTTTACTGACATGGTGAATAATTGCA 1806  
Db 389 ArgValAlaArgIleAlaLysThrAlaGluAspIleLeuProGlu-----ArgIleTyr 406  
QY 1807 CTCTGTCTGCCAAATATTTGCAACCCAGCCCAATACCACCTCCTGAAACAGAGATGAATCAT 1866  
Db 407 ArgLeuValGlyAspVal-----Glu 413  
QY 1867 TCTGTACGATGTCAACAGGACAGATCGCCAGTCTTTTAGCTTAATGCTTTCTTCTGCACA 1926  
Db 414 SerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431  
QY 1927 TTTCCTCCGACGGAAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTC 1986

QY 904 GAGCCTGAGTCACCAATCGATGTA----- 927  
Db 128 AsnAsnLysSerAsnIleAspValAlaIleAsnSerAspGluAspGluLeuValLeu 147  
QY 928 ---GACAACTCGAGAAACAGTTGTCAAGATTCAGAAAGCAGATGAAGAAACAAGTCAGTC 984  
Db 148 GluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnLysLeuSerGlnAspLeu 167  
QY 985 TTT---GATGAGCAAGAT----- 999  
Db 168 PheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrGln 187  
QY 1000 -----GATCGTTCCTCCCAACAGCAAAATAAATACTTTCAAGTTGCCAAGCAAGA 1047  
Db 188 LeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThr 207  
QY 1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGAAGTGAAGTTAGATTGCAT 1107  
Db 208 GluAlaAspSer----- 211  
QY 1108 TTCCAATTTGAAGGAGAAATAATAGTCTGGGACCAGTGACTTAATGCCAAGCCATCTGGA 1167  
Db 212 ---ThrPheValGlyGluAspSerLysAlaThrLysThrValArgThrSerSerSer 230  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAGTTCCAAAGCAGCATGGAAAAGG----- 1221  
Db 231 PheLeuSerThrValSerThr---CysGluAlaProAlaLysGlyArgAlaArgMetTyr 249  
QY 1222 GATTCATAAATTACAGATCATTTTCATGAGAAATTTCCAAAGTCAGAG----- 1266  
Db 250 GlnLysGluLeuGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnPro 269  
QY 1267 -----GACAGAAAGAAAGAAACAATGTGAAGTCAGATCAAAAGA 1305  
Db 270 AspLeuAsnLysValAspProAspArgAsnTyrArgTyrCysThrIle----- 285  
QY 1306 ACAGAAAGGAAGATTCCAAAATACATCCCACTTAACCTCCCTCCAGAGAAAGTGGCTG 1365  
Db 286 -----ProAsnPheProAlaSerGln----- 292  
QY 1366 GGAACCTCTATTGAGGAAATGAGAAAAATGCTCGGTGGTGGATCCATTGCTTCTCTTA 1425  
Db 293 GlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeu----- 307  
QY 1426 AGACCATCTGCAAGTCAACAGTGCAGTGTTCGGGTAGACCTTCTGAGAGCAGGAGGTT 1485  
Db 307 ----- 307  
QY 1486 CCGAAACCTTTTCCAAACACATTACAAAGANTTGTGGGATAACAAACATGTGAAAATGCCT 1545  
Db 308 -----ProGlnArgTyrArgGluPhe----- 314  
QY 1546 TGTTCCGGAACAAAACCTTGTAACCTGTGGGAAGATGAGAATGGTGAGCGA----- 1593  
Db 315 -----AspSerArgGlyArgArgAspSerTyr 324  
QY 1594 -----ACTGCAGGG----- 1602  
Db 325 PheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPhe 344  
QY 1603 -----AGTAGTGGGAGCTCATTCAGACTGCACCTTCTCAACAAA 1641  
Db 345 MetPheValGlyLeuLeuHisAsnMetTyrGluPhe----- 356  
QY 1642 TTCACACGACCCAGAACTTGAAGGATGCGATTCTGAAATACAAATAGTGGCA----- 1692  
Db 357 -----AspProAspIleThrTyrLysLeuProAlaLeuGlu 368  
QY 1693 -----TATTCTAAGAAATGGACCTTTACAGCTTTTGGTTGATTCTGGGATAAGGTACTT 1746  
Db 369 MetTyrTyrLysGluMetSerGluLeuValGlyArgGluGluValLeuGluLysPheAla 388  
QY 1747 GAAGAAGCAGAGGCCCAACATTTATATCATGTCATTTTACCTGACATGGTGAAAATTGCA 1806

Db 389 ArgValAlaAlaArgIleAlaLysThrAlaGluAspIleLeuProGlu-----ArgIleTyr 406  
QY 1807 CTCTGTCTGCCAAATATTTGCACCCAGCCCAATACCACTCCTGAAACAGAGATGAATCAT 1866  
Db 407 ArgLeuValGlyAspVal-----Glu 413  
QY 1867 TCTGTACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACA 1926  
Db 414 SerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431  
QY 1927 TTTCCCGACGGAATGCCAAGATGAAATCGAGTATTTCTAGTTACCCAGACATTAACCTTC 1986  
Db 432 -----AlaArgProAspSerProPheSer-----Phe 440  
QY 1987 AATCGGTTGTTTGAAGGACGTTTCATCAAGGAACACAGAAAAAACTGAAAACTCTTCTGC 2046  
Db 441 CysArgIleLeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThr 460  
QY 2047 TACTTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTT-----ACAAGA 2100  
Db 461 TyrPheAspLysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLys 480  
QY 2101 CAGAGTCTTGAAGATTTT---CCAGAATGG---GAAAGGTGTGAAAAGCCTCTGACACGC 2154  
Db 481 MetAspLysAspThrPheAsnGluGluTyrLysAspLysLysLeuArgSerLeuProGlu 500  
QY 2155 TTACACGTCACCTACGAGGGTACCATAGAAAGCAACGCGGAGCATGCTACAGGTGGAT 2214  
Db 501 ValGluPhePheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAsp 519  
QY 2215 TTTGCAAAATCGTTTGTGGAGGTGGTGACTGGTGGGAGCTTGTACAAAGAAAGAAATC 2274  
Db 520 PheAlaAsnGluHisLeuGlyGlyValLeuAsnHisGlySerValGlnGluGluIle 539  
QY 2275 AGATTTTAAATCAATCCTGAATTGATTGTTTACGGCTGTTCACTGAGGTGCTGGATCAC 2334  
Db 540 ArgPheLeuMetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGln 559  
QY 2335 AATGAGTGTCTTATTATCACAGGTACTGAAACAGTACAGTGAATACAGGCTATGCTGAA 2394  
Db 560 LeuGluAlaIleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHis 579  
QY 2395 ACTTATCGTTGGGCC-----CGAAGCCATGAAGATGGAGTGAA---AAG 2436  
Db 580 ThrLeuLysTyrAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArg 599  
QY 2437 GACGATTGGCAGCGGCTGCACGGAGATCGTTGCCATTCACGCACTTCACCTCAGACGC 2496  
Db 600 AspArgPheGlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGly 619  
QY 2497 -----TACCTCGATCAGTTTGTGCCTGAGAAAAGTGAGACGTGAGCTTAAC 2541  
Db 620 SerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLys 639  
QY 2542 AAGGCTTACTGCGGATTCTCCCTCGCTCGAGTTCCTTCTGAAAATCTTTCTGAGTGGCC 2601  
Db 640 LysAlaSerIleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleVal 658  
QY 2602 ACGGAAAACCTGGGCTGTGGTGCCTTTGGGGGTGACCGCTAGATTAAAGCCTTAATACAG 2661  
Db 659 ThrGlyTyrTyrGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGln 678  
QY 2662 ATCTTGGCAGCTGCTGGCGCTGAACGTGACGTGGTGTATTATTCACCTTTGGGGACTCAGAG 2721  
Db 679 ValIleAlaAlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGlu 698  
QY 2722 TTGATGAGAGACATTTTACAGCATGCACACTTTTCCCTTACGAGAGGAAGCTGGATGTGGA 2781  
Db 699 LeuAlaAlaLysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGly 718  
QY 2782 AAA 2784

Db 345 MetPheValGlyLeuLeuHisAsnMetTrpGluPhe----- 356  
QY 1642 TTCACACGACCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGCA----- 1692  
Db 357 -----AspProAspIleThrTyrLysLeuProAlaLeuGlu 368  
QY 1693 -----TATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTT 1746  
Db 369 MetTyrTyrLysGluMetSerGluLeuValGlyArgGluGluValLeuGluLysPheAla 388  
QY 1747 GAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGCATGCGTGGTGAATTCGA 1806  
Db 389 ArgValAlaAlaArgIleAlaLysThrAlaGluAspIleLeuProGlu-----ArgIleTyr 406  
QY 1807 CTCTGTCTGCCAAATATTTCACCCAGCCCAATACCCTCCTGAAACAGAGATGAATCAT 1866  
Db 407 ArgLeuValGlyAspVal-----Glu 413  
QY 1867 TCTGTACAGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTCTTCTGCACA 1926  
Db 414 SerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431  
QY 1927 TTTCCCGACGGAATGCCAAGATGAAATCGAGTATTCTAGTTACCCAGACATTAACCTTC 1986  
Db 432 -----AlaArgProAspSerProPheSer-----Phe 440  
QY 1987 AATCGTTTGTGAAGGACGTTTCATCAAGGAACACAGAAACCACTGAAACACTCTTCTGC 2046  
Db 441 CysArgIleLeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThr 460  
QY 2047 TACTTTTGAAGAGTCACAGAGAAACCACTACAGGATGGTGACATTT-----ACAAGA 2100  
Db 461 TyrPheAspLysMetSerMetAspProAspGlyAlaValSerPheArgLeuThrLys 480  
QY 2101 CAGAGTCTGAAGATTTT---CCAGATGG---GAAAGGTGTGAAAGCCTCTGCACGCG 2154  
Db 481 MetAspLysAspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGlu 500  
QY 2155 TTACACGTCACCTACGAGGGTACCATAGAGGCAACGGCCGAGGCGTCTACAGGTGGAT 2214  
Db 501 ValGluPhePheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAsp 519  
QY 2215 TTTGCAATCGTTTGTGGAGGTGGTGTGACTGGTGGGACTTGTACAAAGAAGAAATC 2274  
Db 520 PheAlaAsnGluHisLeuGlyGlyValLeuAsnHisGlySerValGlnGluIle 539  
QY 2275 AGATTTTAAATCAATCCTGAATGATTGTTTACGGCTGTTCACTGAGGTGCTGATCAC 2334  
Db 540 ArgPheLeuMetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGln 559  
QY 2335 AATGAGTGTCTTATTATCAGAGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAA 2394  
Db 560 LeuGluAlaIleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHis 579  
QY 2395 ACTTATCGTTGGCC-----CGAAGCCATGAAGTGGAGTGAA---AAG 2436  
Db 580 ThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArg 599  
QY 2437 GACGATTGGCAGCGGCTGCACGGAGATCGTTGCCATTGACGCACTTCACCTTCAGACGC 2496  
Db 600 AspArgPheGlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGly 619  
QY 2497 -----TACCTCGATCAGTTTGTGCTGAGAAAGTGAGACGTGAGCTTAAC 2541  
Db 620 SerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLys 639  
QY 2542 AAGGCTTACTCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAATCTTTCTGCAGTGGCC 2601  
Db 640 LysAlaSerIleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleVal 658  
QY 2602 ACGGGAACCTGGGCTGTGGTGCCTTTGGGGTGACGCTAGATTAAAGCCTTAATACAG 2661

Db 659 ThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGln 678  
QY 2662 ATCTGCGCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTACCTTTGGGACTCAGAG 2721  
Db 679 ValIleAlaAlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGlu 698  
QY 2722 TTGATGAGAGACATTTACAGCATGTCACACTTTCCTTACCAGAGGAAGCTGGATGTTGA 2781  
Db 699 LeuAlaAlaLysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGly 718  
QY 2782 AAA 2784  
Db 719 Lys 719  
RESULT 14  
US-09-511-477-10  
; Sequence 10, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-477-10  
Alignment Scores:  
Pred. No.: 1.27e-33 Length: 726  
Score: 443.50 Matches: 184  
Percent Similarity: 36.62% Conservative: 124  
Best Local Similarity: 21.88% Mismatches: 278  
Query Match: 6.55% Indels: 255  
DB: 4 Gaps: 32  
US-09-302-812-5 (1-3814) x US-09-511-477-10 (1-726)  
QY 622 GAAGAAGACAGAGACAATCAGCAGTTCTTACACCTATAAACTTGCAATAACAAGCCA 681  
Db 14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32  
QY 682 ACAGTAGGAGATGGGCGAGCCAGAGCAACTGTAACTGCAGTGGATCTGCCAGTCTGTG 741  
Db 33 ThrMet-----LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49  
QY 742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACAGAGAGTCTTGTCA 801  
Db 50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSerSer 67  
QY 802 GATGTTGGTCCGAGGACATGGAACCTGGACCAAAA----- 837  
Db 68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87  
QY 838 -----AATGACAACAATG 852  
Db 88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107  
QY 853 ACTGGACAAGAACGCGCTAGGTGAT-----TCGCTCCATTTGAGAAAGAAAGT 903  
Db 108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127



Db 492 Gly-----ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 503  
QY 2629 GGGGGTGACGCTAGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTCGGCTGAACGT 2688  
Db 504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523  
QY 2689 GACGTGTTTATTTCACCTTTGGGACTCAGAGTTGATGAGACACATTTACAGCATGCAC 2748  
Db 524 ProLeuAlaTyrThrPheGlyAsnValGluPheArgAspPheHisGluMetTrp 543  
QY 2749 ACTTTCCTTACCGAGAGGAAGCTGGATGTTGGAAAAGTGTAAGTTATTGCTTAGATAC 2808  
Db 544 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSer 562  
QY 2809 TACAATGAAGAATGCAGAAACTGTTCCACCCTCGACCA-----GAC 2850  
Db 563 TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582  
QY 2851 ATCAAGCTTTATCCATTTCATATACCATGCTGTTGAGTCAAGTGCAGACCACTGACATG 2910  
Db 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599  
QY 2911 CCAGGACAGAAAGCA 2925  
Db 600 ProGlyGluGlyAla 604

RESULT 13  
US-09-302-812-10  
; Sequence 10, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-302-812-10

Alignment Scores:  
Pred. No.: 1.27e-33 Length: 726  
Score: 443.50 Matches: 184  
Percent Similarity: 36.62% Conservative: 124  
Best Local Similarity: 21.88% Mismatches: 278  
Query Match: 6.55% Indels: 255  
DB: 4 Gaps: 32

US-09-302-812-5 (1-3814) x US-09-302-812-10 (1-726)

QY 622 GAAGAAGACAGACAGACAAATCAGCAGTTTCTTACACCTATAAACTTGCAATACAAAGCCA 681  
Db 14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32  
QY 682 ACAGTAGGAGATGGGCGAGCCAGCAAGCAACTGTAACTGCAGTGGATCTCGCCAGTCTGTG 741  
Db 33 ThrMet-----LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49  
QY 742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCTTTGTCA 801  
Db 50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSerSer 67

QY 802 GATGTTGGTCCGAGGACATTGGAACCTGGACCAAAA----- 837  
Db 68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87  
QY 838 -----AATGACAAACAAATTG 852  
Db 88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107  
QY 853 ACTGGACAAGAAAGCAGCCTAGGTGAT-----TCGCCTCCATTTGAGAAAAGAAAGT 903  
Db 108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127  
QY 904 GAGCTGAGTCACCAATGGATGTA----- 927  
Db 128 AsnAsnLysSerAsnIleAspValAlaIleAsnSerAspGluAspGluLeuValLeu 147  
QY 928 ---GACAACTCGAGAAACACTTGTCAAGATTTCAGAGCAGATGAAGAAACAAGTCCAGTC 984  
Db 148 GluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnLeuSerGlnAspLeu 167  
QY 985 TTT---GATGAGCAAGAT----- 999  
Db 168 PheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrThrGln 187  
QY 1000 -----GATCGTTCCTCCAAACAGCAAAATAAACTTCAAGTTGCCAAGCAAGA 1047  
Db 188 LeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThr 207  
QY 1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107  
Db 208 GluAlaAspSer----- 211  
QY 1108 TTCCAATTGAAGGAGAAATAATGCTGGGACCAGTGACTTAATGCCAAGCCATCTGGA 1167  
Db 212 ---ThrPheValGlyGluAspSerLysAlaThrLysThrValArgThrSerSerSerSer 230  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAGTTCCAAAGCAGCATGGAAAAGG----- 1221  
Db 231 PheLeuSerThrValSerThr---CysGluAlaProAlaLysGlyArgAlaArgMetTyr 249  
QY 1222 GATTCTAAAATTACAGATCATTTTCATGAGAAATTTCCAAAGTCAGAG----- 1266  
Db 250 GlnLysGluLeuGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnPro 269  
QY 1267 -----GACAGAAGAAAAGAAACAATGTGAAGTCAGACATCAAGA 1305  
Db 270 AspLeuAsnLysValAspProAspArgAsnTyrArgTyrCysThrIle----- 285  
QY 1306 ACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCAGAGAAAGAGTGGCTG 1365  
Db 286 -----ProAsnPheProAlaSerGln----- 292  
QY 1366 GGAACCTCTATTGAGGAAATGAGAAAAATGCTCGGTGGGTGATCCATTTGCCTTCCTTA 1425  
Db 293 GlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeu----- 307  
QY 1426 AGACCATCTGCAAGTCACACAGTGAAGTGTTCGGGTAGACCTTCTGAGAGCAGAGGTT 1485  
Db 307 ----- 307  
QY 1486 CCGAAACCTTTTCCAACACATTACAAAGATTGTGGGATAACAAACATGTGAAAATGCCT 1545  
Db 308 -----ProGlnArgTrpArgGluPhe----- 314  
QY 1546 TGTTCCGAACAAAACCTGTACCCCTGTGGAAGATGAGAATGGTGAGCGA----- 1593  
Db 315 -----AspSerArgGlyArgArgAspSerTyr 324  
QY 1594 -----ACTGCAGGG----- 1602  
Db 325 PheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPhe 344  
QY 1603 -----AGTAGTGGGAGCTCATTCAGACTGCACACTTCTCAACAAA 1641

QY 2749 ACTTCTTACCGAGGAGCTGGATGTTGGAAAAGTGTACAAAGTTATTGCTTAGATAC 2808  
Db 544 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTyrSer---IleLeuArgSer 562  
QY 2809 TACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA-----GAC 2850  
Db 563 TyrSerArgLeuIleLysGluLysSerLysGluProArgGluAsnLysAlaSerLys 582  
QY 2851 ATCAAGCTTATCCATTATCATATACCATGCTGTTGAGTCAAGTGCAGAGACCACCTGACATG 2910  
Db 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599  
QY 2911 CCAGGACAGAAAGGCA 2925  
Db 600 ProGlyGluGlyAla 604  
RESULT 12  
US-09-511-507-8  
; Sequence 8, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-507-8  
Alignment Scores:  
Pred. No.: 1.49e-94 Length: 768  
Score: 1074.50 Matches: 244  
Percent Similarity: 58.12% Conservative: 96  
Best Local Similarity: 41.71% Mismatches: 194  
Query Match: 15.86% Indels: 51  
DB: 4 Gaps: 15  
US-09-302-812-5 (1-3814) x US-09-511-507-8 (1-768)  
QY 1249 AGAATTTCCAAAGTCAGAGGACAGAAAGAAAGCAATGTGAAGTCAGACATCAAGAACA 1308  
Db 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu----- 62  
QY 1309 GAAAGGAAGATTCCAAATACATCCACCTAACCTCCCTCCA-----GAGAAGAAGTGG 1362  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspAspSerTyr 74  
QY 1363 CTGGGAACCTCTATTGAGGAATGAGAAAATGCCTCGGTGGGTAGACCTTCCTTCC 1422  
Db 75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93  
QY 1423 TTAAGACCA---TCTGCAAGTCACACAGTGAAGTGTGGGTAGACCTTCGAGAGCAGGA 1479  
Db 94 LeuProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGlu 113  
QY 1480 GAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTGTGGGATAACAACATGTGAAA 1539  
Db 114 ThrProProArgProTyrLysSerProGlyLys-----TyrAspSerGluHisValArg 131  
QY 1540 ATGCCTTGTTCGGAACAAAACCTTGATCCCTGTGGAAGATGAGATGGTGAGCGAACTGCA 1599

Db 132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIle 151  
QY 1600 GGGAGTAGGTGGGAGCTCAATCAGACTGCACCTTCTCAACAAATTCACAGACCCCGAAC 1659  
Db 152 AspPheArgTyrGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171  
QY 1660 TTGAAGGATCGATTCTGAAATACAAATGTGGCATAATTCTAAGAAATGGGACTTTACAGCT 1719  
Db 172 LeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTyrHisPheArgAla 191  
QY 1720 TTGGTTGATTCTGGGATAAGTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCC 1779  
Db 192 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211  
QY 1780 ATTTTACCTGACATGGTGAATATGCACTCTGTCTGCCAATATTTGCACCCAGCCATA 1839  
Db 212 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerProVal 231  
QY 1840 CCACTCTCTGAAACAGAAAGATGAATCATCTGTCTGTCCCAATATTTGCACCCAGCCAGT 1899  
Db 232 ProLeuLeuLysHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSerCys 251  
QY 1900 CTTTGTAGTAATGCTTCTTCTGTGCACATTTCCCGACGGAATGCC---AAGATGAAATCG 1956  
Db 252 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSer 271  
QY 1957 GAGTATTTAGTATACCCAGACATAAATTAATTCGTTTGTGAAAGGACGTTTCATCAAGG 2016  
Db 272 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291  
QY 2017 AAACCCAGAAAACTGAAACACTCTTCTGTCTACTTTCGAAGAGTC-----ACAGAGAAA 2070  
Db 292 ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 311  
QY 2071 AAA-----CCTACAGGATGGTGACATTTTACAAGACAGAGT-----CTTGAA 2112  
Db 312 AspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGlu 331  
QY 2113 GATTTTCCAGAAATGGAAAGGTGTGAAAAGCCTCTG-----ACACGCTTACACGTCAC 2166  
Db 332 HisLeuIleAspTyrSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351  
QY 2167 TACGAGGTACCATAGAGGCAACGGCCGAGGCATGCTACAGTGGATTTTGCAATCGT 2226  
Db 352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371  
QY 2227 TTTGTTGGAGGTGGTGTGACTGGTGGGGACTTGTACAAAGAAATCAGATTTTAAATC 2286  
Db 372 TyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheValIle 391  
QY 2287 AATCCTGATGATTGTTTTCACGGCTGTTTCACTGAGGTGCTGGATCACAATGAGTGTCT 2346  
Db 392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgPropheGluAlaLeu 411  
QY 2347 ATTATCAGAGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGG 2406  
Db 412 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTyr 431  
QY 2407 GCCCGAAGCCATGAAGATGGAGTGAAGAAAGCAGATTGGCAGCGCGCTGCACGGAGATC 2466  
Db 432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 451  
QY 2467 GTTGCCATTGACGCACTTCACTTCAAGCAGCTACCTCGATCAGTTGTGCTGAGAAAGTG 2526  
Db 452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet 471  
QY 2527 AGACGTGAGCTTAACAAGCTTACTGCGGATTCTCCCGT-----CCT 2568  
Db 472 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTyrMetValThrProPro 491  
QY 2569 GGAGTTCTCTTGAAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTT 2628

QY 2851 ATCAAGCTTATCCATTATACCATGCTGTGTTGAGTCAAGTGCAGACACCACTGACATG 2910  
|||||  
Db 583 LysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599  
|||  
QY 2911 CCAGGACAGAAGGCA 2925  
|||||  
Db 600 ProGlyGluGlyAla 604  
|||||  
RESULT 11  
US-09-511-477-8  
; Sequence 8, Application US/095111477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-477-8

Alignment Scores:  
Pred. No.: 1.49e-94 Length: 768  
Score: 1074.50 Matches: 244  
Percent Similarity: 58.12% Conservative: 96  
Best Local Similarity: 41.71% Mismatches: 194  
Query Match: 15.86% Indels: 51  
DB: 4 Gaps: 15

US-09-302-812-5 (1-3814) x US-09-511-477-8 (1-768)

QY 1249 AGAATTTCAAGTCAGAGGACAGAGAAAGAAACAATGTGAAGTCAGACATCAAGAACA 1308  
|||||  
Db 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGlu----- 62  
|||  
QY 1309 GAAAGGAAGATTCCAAAATACATCCCACTTAACCTCCCTCCA-----GAGAAGAAAGTGG 1362  
|||||  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74  
|||  
QY 1363 CTGGGAACCTCTATTGAGGAAATGAGAAATGCCTCGGTGGGATCCATTGCCTTCC 1422  
|||  
Db 75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93  
|||  
QY 1423 TTAAGACCA---TCTGCAAGTCACACAGTGAAGTGTTCGGGTAGACCTTCTGAGAGCAGGA 1479  
|||  
Db 94 LeuProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGlu 113  
|||  
QY 1480 GAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTGTGGGATAACAACATGTGAA 1539  
|||||  
Db 114 ThrProProArgProTyrLysSerProGlyLys-----TrpAspSerGluHisValArg 131  
|||  
QY 1540 ATGCCTTGTTCGGAACAAACTTGTACCTGTGGAAGATGAGAATGTTGAGCGAACTGCA 1599  
|||||  
Db 132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrIle 151  
|||  
QY 1600 GGGAGTAGTGGAGCTCATTCAGACTGCACTTCTCAACAAATTTCACACGACCCCAAG 1659  
|||||  
Db 152 AspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171  
|||  
QY 1660 TTGAAGGATGCGAATCTGAAATACAAATGTGGCATATTCTAAGAAATGGGACTTTACAGCT 1719  
|||||

Db 172 LeuGlnAlaAlaIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAla 191  
QY 1720 TTGGTTGATTTCTGGGATAAGGTACTTTGAAGAAGCAGAGGCCCAACATTATATCAGTCC 1779  
|||  
Db 192 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211  
|||  
QY 1780 ATTTTACCTGACATGGTGAATAATGCACACTCTGTCTGCCAAATATTGCAACCCAGCCAATA 1839  
:::|  
Db 212 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerProVal 231  
|||  
QY 1840 CCACCTCCTGAAACAGAGATGAATCATCTGTCTCAGCATGTCACAGGAACAGATCGCCAGT 1899  
|||  
Db 232 ProLeuLeuLysHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSerCys 251  
|||  
QY 1900 CTTTGTAGCTAATGCTTTCTTCTGACATTTCCCGACGGAATGCC---AAGATGAAATCG 1956  
|||  
Db 252 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSer 271  
|||  
QY 1957 GAGTATTTAGTTACCCAGACATTAATCAATCGTTGTTGTTGAGGACGTTTCATCAAGG 2016  
|||  
Db 272 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291  
|||  
QY 2017 AAACAGAGAAACTGAAACACTCTTCTGCTACTTTTCGAAGATC-----ACAGAGAAA 2070  
|||  
Db 292 ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 311  
|||  
QY 2071 AAA-----CCTACAGGATGGTGACATTTACAAGACAGAGT-----CTTGAA 2112  
|||  
Db 312 AspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGlu 331  
|||  
QY 2113 GATTTTCCAGAATGGGAAAGGTGTGAAAGCCTCTG-----ACACGCTTACACGTCAC 2166  
:::|  
Db 332 HisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351  
|||  
QY 2167 TACGAGGTTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTGCAATCGT 2226  
|||  
Db 352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371  
|||  
QY 2227 TTTGTTGAGGTGGTGTGACTGGTGGCGGACTTGTACAAGAAAGAAATCAGATTTTAATC 2286  
:::|  
Db 372 TyrLeuGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheValIle 391  
|||  
QY 2287 AATCCTGAATGATGTTTTCACGGCTGTTCACGTGCTGAGGTGCTGGATCAATGAGTGTCT 2346  
|||  
Db 392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 411  
|||  
QY 2347 ATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGG 2406  
:::|  
Db 412 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTrp 431  
|||  
QY 2407 GCGCGAAGCCATGAAGATGGGAGTGAAGAGGACGATTGGCAGCGCGCTGCACGGAGATC 2466  
:::|  
Db 432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 451  
|||  
QY 2467 GTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCTCGCTGAGAAAGTG 2526  
|||  
Db 452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisGlnTyrArgGluAspLeuMet 471  
|||  
QY 2527 AGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGT-----CCT 2568  
|||  
Db 472 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 491  
|||  
QY 2569 GGAGTTCCTTCTGAAATCTTCTGAGTGGCCACGCGGAAACTGGGGCTGTGGTGCCTTT 2628  
|||  
Db 492 Gly-----ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 503  
|||  
QY 2629 GGGGTGACGCTAGATTAAAGCCTTAATACAGATCCTGGCAGCTGCTGGGCTGAACGT 2688  
|||  
Db 504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523  
|||  
QY 2689 GACGTGGTTTATTTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCAC 2748  
:::|  
Db 524 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMetTrp 543  
|||







|    |      |           |        |        |        |        |        |        |        |        |        |        |      |      |      |
|----|------|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|------|------|------|
| QY | 787  | GAGAGTCC  | TTTTGT | CAGATG | TTGGT  | CCGAGG | ACAT   | TGGA   | CTGG   | ACCA   | AAAA   | ATG    | CACA | C    | 846  |
| Db | 261  | GluSerPro | LeuSer | AspThr | GlySer | GluAsp | ValGly | ThrGly | LeuLys | AsnAla | Asn    |        |      |      | 280  |
| QY | 847  | AAATTG    | ACTG   | CAAG   | AAAG   | CAGC   | CTAG   | GTG    | ATTC   | GCCT   | CCAT   | TTG    | AGAA | AGAA | 906  |
| Db | 281  | ArgLeu    | AsnArg | GlnGlu | SerSer | LeuGly | AsnSer | ProPro | PheGlu | LysGlu | SerGlu |        |      |      | 300  |
| QY | 907  | CCTG      | AGTCA  | CAAT   | GGAT   | GTAG   | CAACT  | CGA    | AAAC   | CAGT   | TGT    | CA     | AGAT | TC   | 966  |
| Db | 301  | ProGlu    | SerPro | MetAsp | ValAsp | AsnSer | LysAsn | SerLys | AsnSer | CysGln | AspSer | GluAla | Asp  |      | 320  |
| QY | 967  | GAAGAA    | CAAG   | TCCAG  | TCTTT  | GATG   | AGCA   | AGAT   | GC     | TGTT   | CTCT   | CC---  | CAAC | AGCA | 1023 |
| Db | 321  | GluGlu    | ThrSer | ProGly | PheAsp | GluGln | GluAsp | SerSer | AlaGln | ThrAla | Asn    |        |      |      | 340  |
| QY | 1024 | AAACTT    | CAAG   | TGTC   | CAAG   | CAAG   | AGAC   | TGAT   | GGCG   | ATCT   | TAGG   | AAAC   | CGT  | ATT  | 1083 |
| Db | 341  | LysPro    | SerArg | PheGln | ProArg | GluAla | AspThr | GluLeu | ArgLys | ArgSer | SerAla |        |      |      | 360  |
| QY | 1084 | AAGG      | AAAG   | TGA    | AGTT   | AGAT   | TGCA   | TTT    | CCA    | ATT    | TGAA   | ---    | GG   | AGAA | 1140 |
| Db | 361  | LysGly    | GlyGlu | IleArg | LeuHis | PheGln | PheGlu | GlyGly | GlySer | ArgAla | GlyMet |        |      |      | 380  |
| QY | 1141 | AGTG      | ACTT   | AAAT   | GCCAA  | GCCAT  | CTCG   | AAAC   | TCTT   | CAGC   | CTT    | TAAT   | GTAG | AGT  | 1200 |
| Db | 381  | AsnAsp    | ValAsn | AlaLys | ArgPro | GlySer | ThrSer | SerSer | LeuAsn | ValGlu | CysArg | Asn    |      |      | 400  |
| QY | 1201 | TCCA      | AGC    | AGC    | ATG    | GA     | AAAG   | GGAT   | TCTA   | AAAT   | TAC    | AGAT   | CAT  | TT   | 1260 |
| Db | 401  | SerLys    | GlnHis | GlyArg | LysAsp | SerLys | IleThr | AspHis | PheMet | ArgVal | ProLys |        |      |      | 420  |
| QY | 1261 | TCAG      | AGG    | CAG    | AGAA   | AGAA   | CAAT   | GTGA   | AGTC   | AGAC   | ATCA   | AAAG   | AAAC | AGAA | 1320 |
| Db | 421  | AlaGlu    | AspLys | ArgLys | GluGln | CysGlu | MetLys | HisGln | ArgThr | GluArg | LysIle |        |      |      | 440  |
| QY | 1321 | CCAA      | AAAT   | ACAT   | CCCC   | ACCT   | ACCT   | CCCT   | CCAG   | AGA    | AGAT   | GGCT   | GGAA | CT   | 1380 |
| Db | 441  | ProLys    | TyrIle | ProPro | HisLeu | SerPro | AspLys | LysTrp | LeuGly | ThrPro | IleGlu |        |      |      | 460  |
| QY | 1381 | GAAAT     | GAGAA  | AAAT   | GCCT   | CGGT   | GCGAT  | CCAT   | TTGC   | CTT    | CCTT   | TAAG   | ACC  | AT   | 1440 |
| Db | 461  | GluMet    | ArgArg | MetPro | ArgCys | GlyIle | ArgLeu | ProPro | LeuArg | ProSer | AlaAsn |        |      |      | 480  |
| QY | 1441 | CAC       | AGT    | GACT   | GTTC   | GGT    | AGAC   | CTT    | CTG    | AGAG   | CAGG   | AGAG   | GTTC | CGAA | 1500 |
| Db | 481  | HisThr    | ValThr | IleArg | ValAsp | LeuLeu | ArgIle | GlyGlu | ValPro | LysPro | PhePro |        |      |      | 500  |
| QY | 1501 | ACAC      | ATT    | TACAA  | AGAT   | TTGT   | GGG    | ATA    | CAAC   | ATGT   | GAAA   | ATGC   | CTT  | GT   | 1560 |
| Db | 501  | ThrHis    | PheLys | AspLeu | TrpAsp | AsnLys | HisVal | LysMet | ProCys | SerGlu | GlnAsn |        |      |      | 520  |
| QY | 1561 | TTGT      | ACC    | CTGT   | GGA    | AGAT   | GGT    | GAG    | CGAA   | CTGC   | AGG    | GAG    | TAG  | GT   | 1620 |
| Db | 521  | LeuTyr    | ProVal | GluAsp | GluAsn | GlyGlu | ArgAla | AlaGly | SerArg | TrpGlu | LeuIle |        |      |      | 540  |
| QY | 1621 | CAG       | ACT    | GCA    | CTT    | CTCA   | CAAA   | ATT    | CAC    | AGC    | ACCC   | CAGAA  | CTT  | GA   | 1680 |
| Db | 541  | GlnThr    | AlaLeu | LeuAsn | ArgLeu | ThrArg | ProGln | AsnLeu | LysAsp | AlaIle | LeuLys |        |      |      | 560  |
| QY | 1681 | TACA      | ATG    | TGG    | CA     | TAT    | TCTA   | AGAAA  | ATGG   | GACT   | TTT    | TAC    | AGCT | TT   | 1740 |
| Db | 561  | TyrAsn    | ValAla | TyrSer | LysLys | TrpAsp | PheThr | AlaLeu | IleAsp | PheTrp | AspLys |        |      |      | 580  |
| QY | 1741 | GTAC      | TTGA   | AGA    | AGC    | AGAG   | CGCC   | CAAC   | ATT    | TAT    | TAT    | CAGT   | CCAT | TT   | 1800 |
| Db | 581  | ValLeu    | GluGlu | AlaGlu | AlaGln | HisLeu | TyrGln | SerIle | LeuPro | AspMet | ValLys |        |      |      | 600  |
| QY | 1801 | ATTG      | CAC    | TCT    |        |        |        |        |        |        |        |        |      |      |      |

|    |      |  |  |  |                                    |      |
|----|------|--|--|--|------------------------------------|------|
| QY | 1861 | AATCATCTCTGT   | CACGATGT   | CACAGGAA                               | CAGATCGCCAGTCTTTTAGCTAATAGCTTTCTTC | 1920 |
| Db | 621  | AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe | 640  |  |                                    |      |
| QY | 1921 | TGCACATTTCCC   | GACGGAATG  | CAAGATGAAATCGGAGTATCTAGTTACCCAGACATT   | 1980                               |      |
| Db | 641  | CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle | 660  |  |                                    |      |
| QY | 1981 | AAC TTCAATCGGTTGTTTGAAGSACGTTT                               | CATCAAGSAAACAGAAAAC                              | CTGAAAACATC                            | 2040                               |      |
| Db | 661  | AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu | 680  |  |                                    |      |
| QY | 2041 | TTCTGCTACTTTT  | CGAAGAGT   | CACAGAGAAAAACCTACAGGATTGGTGACATTACAAGA | 2100                               |      |
| Db | 681  | PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg | 700  |  |                                    |      |
| QY | 2101 | CAGAGTCTTGAAGATTTT   | CCAGAATGGGAAAGGTG                                | GAAAAAGCCTCTGACACGCTTACAC              | 2160                               |      |
| Db | 701  | GlnSerLeuGluAspPheProGluTyrGluArgCysGluLysLeuLeuThrArgLeuHis | 720  |  |                                    |      |
| QY | 2161 | GTCACTTACGAGGGT  | TACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCA   | 2220                                   |                                    |      |
| Db | 721  | ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla | 740  |  |                                    |      |
| QY | 2221 | AATCGTTTTTGTGGAGGTGCTGTGACTGGTGC                             | GGGACTTGTACAAGAGAAATCAGATTT                      | 2280                                   |                                    |      |
| Db | 741  | AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe       | 760  |  |                                    |      |
| QY | 2281 | TTAATCAATCCTGAATGAATGTTT                                     | CACGGCTGTCTACTGAGGTGCTGCATCAATGAG                | 2340                                   |                                    |      |
| Db | 761  | LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu | 780  |  |                                    |      |
| QY | 2341 | TGCTCTTATTATCACAGGTACTGAACAGTAC                              | AGTGAATACACAGGCTATGCTGAAACTTAT                   | 2400                                   |                                    |      |
| Db | 781  | CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr | 800  |  |                                    |      |
| QY | 2401 | CGTTGGCCCGAAGCCATGAAGATGGGAGTGA                              | AAAGGACGATTGGCAGCGCGCTGCACG                      | 2460                                   |                                    |      |
| Db | 801  | ArgTrpAlaArgSerHisGluAspArgSerGlu                            | ArgSerGluArgAspTrpGlnArgArgThrThr                | 820                                    |                                    |      |
| QY | 2461 | GAGATCGTTGCCATTGACGCACCTT                                    | CACCTCAGACGCTACCTCGATCAGTTGTGCCTGAG              | 2520                                   |                                    |      |
| Db | 821  | GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu | 840  |  |                                    |      |
| QY | 2521 | AAAGTGAGACGTGAGCTTAA   | CAAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCTCTTCT        | 2580                                   |                                    |      |
| Db | 841  | LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer | 860  |  |                                    |      |
| QY | 2581 | GAAAAATCTTCTG  | CAGTGGCCACGGGAAACTGGGGCTGTGGTGCCCTTTGGGGGTGACGCT | 2640                                   |                                    |      |
| Db | 861  | GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla | 880  |  |                                    |      |
| QY | 2641 | AGATTAAAGCCTTAATACAGATCCT                                    | TGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTTAT             | 2700                                   |                                    |      |
| Db | 881  | ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr    | 900  |  |                                    |      |
| QY | 2701 | TTCACTTTGGGGACTCAGAGTTGATGAGAGACATTT                         | TACAGCATGCACACTTTCCTTACC                         | 2760                                   |                                    |      |
| Db | 901  | PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr | 920  |  |                                    |      |
| QY | 2761 | GAGAGGAAGCTGCATGTTGGAAAAGTGTA                                | CAAGTTATTGCTTAGATACATAATGAAGA                    | 2820                                   |                                    |      |
| Db | 921  | GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu | 940  |  |                                    |      |
| QY | 2821 | TGCAGAAACTGTTCCACCCCTGGACCAGACATCA                           | AGCTTTATCCATTATATACCATGCT                        | 2880                                   |                                    |      |
| Db | 941  | CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla | 960  |  |                                    |      |
| QY | 2881 | GTTGAGTCAACTGCAGAGACC  | ACTGACATGCCAGGACAGAGCGCAGGC                      | 2928                                   |                                    |      |
| Db | 961  | ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly             | 976  |  |                                    |      |

## RESULT 10